1. Executive Summary

Pearson is pleased to respond to the Partnership for Assessment of Readiness for College and Careers (PARCC) Operational Assessments RFP #40-000-13-00027 issued by the New Mexico Public Education Department. With the support of our subcontractors: Caveon, ETS, Measured Progress, and WestEd, we will develop and implement the PARCC Performance-Based Assessments and End-of-Year/End-of-Course Assessments.

The PARCC Operational Assessments Team

Recognizing PARCC's need for a solution that includes strong collaboration among organizations, and that can provide expertise, scalability and high-quality solutions in developing the assessment system, we propose the following division of labor between Pearson and our subcontractors.

Expert Team for PARCC Assessments	
Component	**Primary Responsibility**
Test Development	Pearson, ETS, WestEd
Assessment Administration	Pearson
Psychometric Services	Pearson, ETS Measured Progress (Quality Control) Caveon (Data Forensics: Internet Monitoring)
Reporting	Pearson
Standard Setting	Pearson ETS, WestEd (Content Facilitators)
Program Management	Pearson

Pearson's subcontractors will provide PARCC with added psychometric, content, and quality assurance expertise. ETS, Measured Progress, Pearson, and WestEd all work together in support of current PARCC contracts. Caveon also works with Pearson in support of a number of state assessment programs. These experiences working together were leveraged in designing solutions for this proposal, as described in the following pages.

Test Development

Pearson has partnered with the best item writing companies across the testing industry. Up to 6,800 items/tasks/texts will be developed each year for ELA/literacy grades 3-11 and mathematics grades 3-8 and high school (Algebra I, Geometry, Algebra II, and Integrated Math I, II and III). This provides PARCC both the capacity and expertise needed to meet these development volumes. We have divided the item development work into logical and coherent pieces among members of the PARCC Operational Assessments team. For

ELA/literacy, the development leads are Pearson for grades 3 to 6, WestEd for grades 7 and 8, and ETS for high school. For mathematics, Pearson will lead grades 3 to 5, and ETS will lead grades 6 to 8 and all high school item development.

We have included sample technology-enhanced items, ELA/literacy items, and mathematics items, that demonstrate our understanding of the item types that need to be created for PARCC. PARCC places a great deal of importance on innovation and with the support of our partners we have also included tasks that push the boundaries of assessment.

Our proposal includes a thorough discussion of the item review process that will be followed by Pearson and our partners, including internal and committee member reviews.

Assessment Administration

Beginning in the 2014–15 school year, the PARCC Summative Assessments in ELA/literacy and mathematics will be available for states to administer to students in grades 3 through 8 and high school, in paper-based and computer-based modes. Students will take two components in both content areas: the Performance Based Assessment (PBA) and the End-of-Year (EOY) Assessment. The assessments will be delivered during one of the administration windows outlined below each school year. Note: the RFP did not include anticipated testing volumes for block windows, only total number of students across the three administrations that should be assumed as the base (8,196,118). Up to 500,000 students may test as part of the Fall/Winter Block 2014, and a similar number for the Spring 2015 Block. The Fall/Winter Block 2014 will be a paper-based administration, beginning in December. Both paper and computer-based testing will be provided for all subsequent administrations.

Administration Window	Approximate Timing	Approximate Number of Students (50% computer-based/ 50% paper-based)
Fall/Winter Block (high school only)	PBA: November-December EOY: December-January	68,985
Traditional Year	PBA: Mid-February-Mid April EOY: Early April-Late May	8,058,148
Spring Block (high school only)	PBA: Early April-Mid May EOY: Early May-Mid June	68,985
Total Number of Students		8,196,118

In creating our Base Cost Proposal, we estimated the number of students that will test during each window as shown in the third column of the table. For determining pricing tiers, one high school test (e.g., Algebra) I is equivalent to ½ student. In grades 3-8 each student counts as one student.

Test Construction

Our team will build both the computer-based and paper-based forms, as well as provide accommodated versions. We will build a set of 10 core forms with a small number of matrix items on each EOY and mathematics PBA form. ELA/literacy PBA items will be included on a standalone form for administration during the EOY operational testing window. For the first year, we will use the anticipated five blueprint sets to build 10 core online forms. In the following years, the form counts for Algebra I, Geometry, and Integrated Mathematics (I, II, and III) will change to 12, 12, and 8, respectively, based on the planned field test counts. See section V.B for further information about online versus paper form assumptions by test each year of the contract.

The following table illustrates the large number of forms that need to be developed each year. This includes both computer-based and paper-based forms as well as all operational standalone field tests, and embedded field test variant forms. We will work with PARCC to develop a detailed schedule for managing the multiple review steps.

Number of Forms Needed Each Year			
Year	Math	ELA	Total
2014-15	2,144	1,762	**3,906**
2015-16	2,116	1,762	**3,878**
2016-17	1,347	1,513	**2,860**
2017-18	1,347	1,493	**2,840**

Paper-Based Testing

Since the majority of printing work will be done in-house, we can control the production environment, press schedule, and quality process for print materials—resulting in quality printed materials for the PARCC Assessments. Early in the process we will provide PARCC with a print materials plan and the Print Specifications for review. Quantities for paper-based materials will be collected through the registration process, entered into PearsonAccess. We recognize that this the first any of these states will be administering PARCC Operational Assessments and we are prepared to provide extensive training materials for school and district staff, including technology staff responsible for online readiness.

Two months before each administration window, we will provide PARCC with packaging, distribution, and retrieval specifications developed under the rigor of our ISO 9001—certified quality management system. We will meet the distinct needs of each state by packaging and distributing materials at the district or school level, whichever the state prefers. To confirm the accuracy of our packaging of each order, we will provide automated quality control verification that accounts for materials in real-time as the distinct barcode on each packaging component is picked and scanned. Once returned boxes arrive at Pearson for processing, the box label will be scanned for accountability and will become part of the daily receipt log. Secure materials that are not returned will be tracked using the missing materials process to safeguard the integrity of the assessment material.

Section V.B Assessment Administration includes extensive details about our scanning capabilities, throughput, and quality processes that will be leveraged to scan the millions of paper-based documents that will be received for PARCC. We have used these processes for other large state assessment and national assessment programs, including the California STAR program, Texas state assessment programs, and NAEP.

Computer-Based Testing

Through this contract we will provide test registration, test delivery, and item banking systems, including hosting, maintenance, management, and administration of these systems.

Scoring

The integrity of the machine and human scoring processes are paramount to providing PARCC rich, accurate field test data (to inform ongoing item selection for operational forms) and exacting operational test data for reporting results of the high-stakes tests. Pearson will monitor all aspects of the scoring procedures, including key-based and rule-based machine scoring and hand scoring for constructed response items and performance tasks.

PARCC will require innovative approaches to scale hand scoring for unprecedented volumes, and to provide consistency and reliability of scoring across items, content areas, and grades. We offer broad experience scoring nationwide assessments—including PARCC items. We have a next generation distributed scoring platform that leverages a nationwide pool of 100,000 scoring experts and that is supported by regional scoring hubs. We have the capacity to scale our scoring services to meet PARCC's needs.

Handscoring costs make up nearly 50 percent of the costs of the PARCC assessments. We are eager to conduct the automated scoring efficacy study later this year, in coordination with ETS and PARCC, with the plan to phase in automated scoring beginning with the spring administration of the first operational year.

Psychometric Services

The psychometric services in support of the PARCC operational assessments are a small but critically important component of the project. We have assembled a team combining the strengths of several subcontractors to deliver accurate, timely, and thoughtful psychometric services.

ETS will serve as the psychometric lead for the contract and have primary responsibility for the deliverables associated with the data analysis of summative, field test, and retest administrations as described in **Section V.C.1** of the RFP. Pearson and ETS will collaborate on the required research studies. Pearson will provide oversight and coordination of the psychometric work. These tasks will include replicating some of the critical psychometric analyses associated with equating and field-test item calibration. Pearson will also be responsible for data forensics and the technology and data solutions. Measured Progress will provide the independent audit and analysis of the psychometric services as required in **Section V.C.7**, and Caveon Test Security will provide services to monitor the internet (and social media) for breaches of test security.

Reporting

The data management and reporting contractor from the Technology Bundle RFP will provide Pearson with the necessary training and documentation needed for Pearson to learn and configure the system. After receiving training on the reporting system, we will work closely with the vendor to understand the final state in which the reports will be turned over and understand the results of the reporting design studies. We will work with PARCC to use the results to make any final revisions needed to produce final report designs.

Standard Setting

The standard-setting process for the PARCC summative assessments will integrate PARCC's College- and Career-Ready (CCR) Determination Policy, policy-level and subject- and grade-specific Performance Level Descriptors (PLDs), empirical data, and content expert judgment to set five performance levels for each assessment. Pearson proposes an Evidence Based Standard Setting (EBSS) process, an approach we pioneered that supports policy claims related to performance standards through systematic research designed to inform the judgments made by content experts. EBSS lends itself well to creating a system of aligned performances standards starting with college and career readiness and linking down from high school to middle school to elementary school. We believe it fits well with PARCC's intended use of empirical studies in the standard setting process to support the policy statements related to student's readiness for college and careers.

Program Management

We have assembled a highly experienced project management team that will support this project in the areas of Operations, Forms Publications, Scoring, Organizational Quality, Software & Technology Services, and Test, Measurement & Research Services, as well as the work of our subcontractors.

Since the founding of our US assessment business in 1962, we have been a trusted partner for both student and educator assessments, including providing test development and delivery services for large-scale K–12 assessment programs in more than 25 states, including several PARCC states: Arizona, Florida, Illinois, Ohio, Maryland, New York, and Tennessee. Many of the key personnel proposed for the PARCC Operational Assessments program have experience working on one or more of these programs and are currently supporting the PARCC Item Development and Assessment Administration programs.

Summary

In summary, with the support of Caveon, ETS, Measured Progress, and West Ed, the Pearson collaborative brings the requisite expertise and capabilities to deliver this project. We will draw upon our widespread experience in working with PARCC and numerous states to develop, administer, score, and report the new PARCC Operational Assessments.

2. Corporate Capabilities

Requirement

2. Corporate Capabilities
This section shall provide a description of the ***Offeror's qualifications and prior*** experience performing tasks similar to those required in this RFP. The discussio***n shall*** include a description of the Offeror's background and relevan***t experience that qualifies it*** to provide the products and services required by this RFP.

Experience: To complete the documentation of corporate ***capabilities, the Offeror must*** document the contracted services for three (3) previo***us projects similar to the one*** described in this RFP or any other projects that document ***its corporate capability. For*** each, the documentation shall include a descriptio***n of the services and products*** delivered; the contract period; and the name, address, and ***telephone number of a contact*** person for each of the contracting entities. T***he New Mexico Public Education*** Department reserves the right to contact the references regarding the services provided.

Disclosure: The documentation of corporate qualification***s must also include disclosure*** statements about all contractual situations occurring within ***the past five years that have*** led to the collection of credits, reimbursements, assessment of penaltie***s and other forms*** of compensation or cancellation of a contract by con***tracting organizations for the*** Offeror's nonperformance. The Offeror shall also prov***ide information related to the*** intent or any knowledge of potential buyout or corporate/business ow***nership changes for*** their company during the periods of services proposed for this RFP.

Response

Qualifications and Prior Experience

The new assessments under development by PARCC are very different from what most states have today. The PARCC assessments require a focus on measuring what students know—a measure of whether or not students are on track to graduate ready for college and careers—and they require providing meaningful and timely results to teachers to guide learning and instruction—results that are comparable across states and across the nation and results that are internationally-benchmarked.

PARCC is very interested in providing a highly efficient and collaborative environment as consortium states implement the new assessments in the 2014-2015 school year and beyond. The consortium encourages organizations to work together in strong partnerships, and to provide high quality, cost effective, and innovative solutions. Pearson and its subcontractors—Caveon, ETS, Measured Progress, and WestEd—have worked together to develop the solutions and this proposal for the PARCC Operational Assessments.

The following is a description of roles and responsibilities, which leverages the strengths of each organization on current PARCC programs.

Expert Team for PARCC Assessments	
Component	**Primary Responsibility**
Test Development	Pearson, ETS, WestEd Measured Progress
Assessment Administration	Pearson
Psychometric Services	Pearson, ETS Measured Progress (Quality Control) Caveon (Data Forensics: Internet Monitoring)
Reporting	Pearson
Standard Setting	Pearson ETS, WestEd (Content Facilitators)
Program Management	Pearson

Team Role Allocation. For the PARCC Partnership, Pearson has assembled a team of experts each bringing a company specialty to the united effort.

Redefining Relationships with PARCC

Through various projects with PARCC and PARCC states, Pearson has demonstrated that we are a solutions-oriented organization, and a true partner in the development and implementation of assessment programs. As programs evolve and requirements change, we offer well-thought-out solutions for our customers.

For the Technology Readiness Tool, this has included maintaining a flexible and iterative approach to the sometimes changing needs on the part of the consortia, coupled with clear communication regarding recommendations or risks and potential impacts to schedule or feature support, where applicable.

For the PARCC Item Development project, we have demonstrated our ability to be agile and provide solution-oriented thinking at various points over the past two years. We have evaluated and implemented a number of scope changes requested by PARCC, including those focusing on accessibility, cognitive complexity, research studies, and changes in the test blueprint. In partnership with ETS we were able to put together a research study proposal, including a detailed timeline and cost estimate, in less than a week. Upon acceptance, we quickly mobilized the additional resources necessary to be able to conduct the cognitive labs and pilots required, within a short amount of time. Both ETS and Pearson worked with PARCC for the Item Development project, of which Phase 2 is now underway.

For the PARCC Assessment Administration project, we have remained flexible throughout the tryout and field test process. Working together with PARCC, Pearson signed Amendment 3 to provide online testing for the Spring 2014 Field Test. Assessing Pearson's readiness for PARCC online testing has been a highly transparent and collaborative process with PARCC.

The level of sharing around areas such as User Acceptance Testing provides PARCC with clear insight about our processes and systems and facilitates a better solution for end-users.

Through our response to a variety of PARCC RFPs, including the Diagnostic Assessments and the Technology Bundle, we have spent countless hours designing solutions with PARCC students, teachers, and parents in mind. We have responded to requirements that are the first of their kind in this industry, with solutions that will lead the way in terms of quality, efficiency, and cost effectiveness. We bring that same power of innovation and thought leadership to this proposal.

Providing New Assessments that Measure and Track Students' Readiness for College and Careers

Over the last two years, we have worked closely with the GED Testing Service to develop the next generation GED, an online assessment, aligned to the Common Core State Standards (CCSS)—ready for use in January 2014. As part of this project, we established efficient, transparent, and reliable test design and development techniques that reflect the principles of evidence-centered design (ECD). Our GED experience, combined with the PARCC ECD experience will further benefit the PARCC program as new items/tasks are developed in 2014–2018.

In addition, we are also working with many states to develop assessments aligned to the CCSS. For example, in June 2011 the Commonwealth of Kentucky awarded Pearson the contract to provide all assessments for Grades 3–8 and On-Demand Writing at High School for grades 10–11. The new paper test (mandated by Senate Bill 1) is called K-PREP (Kentucky Performance Rating for Educational Progress). K-PREP is a large-scale assessment designed to measure student knowledge and skills against pre-determined Kentucky Core Academic Standards in the content areas of Reading, Math, Writing, Science, and Social Studies. The test consists of multiple-choice, short answer and extended response items. For the spring 2013 assessment, we developed 8-14 forms depending on the grade for grades 3–8 for a total of 66 forms. Each form covers all content areas. For grades 10 and 11 we developed one form per grade for the writing component.

Pearson currently supports item and test development for large-scale assessments in 16 states, including Puerto Rico. Whether or not states have adopted the CCSS, every state is interested in improving its assessment program to measure and track students' readiness for college and careers. Pearson has had the opportunity to contribute to this discussion with many states, as well as for national assessment programs such as PARCC.

Quality—It is the Foundation for Everything We Do

The foundation of our business is a sound quality management system. ISO offers the comprehensive quality framework necessary to demonstrate adherence to best practices and the establishment of a true culture of continuous quality improvement throughout our organization. Pearson led the adoption of ISO standardization and certification in the assessment industry. We have the most capability in the assessment industry, with assessment operations spanning four campuses—in Cedar Rapids and Iowa City, Iowa, Owatonna, Minnesota, and Austin, Texas—with all operations certified to the ISO 9001:2008 standard. Every employee in printing, packaging, distribution, data preparation, scanning and performance scoring is required to adhere to these strict standards.

Strong Partnerships with Subcontractors

PARCC encourages responses that demonstrate strong partnerships between organizations. Pearson, ETS, Caveon, Measured Progress, and WestEd have worked collaboratively together on several state and national assessment programs with a division of labor similar to what is proposed for the PARCC Operational Assessments. ETS, Caveon, Measured Progress, and WestEd provide Pearson and PARCC with added psychometric, content, and quality assurance expertise—together we can pave the way for a successful next generation assessment program.

Pearson has a proven record of effectively managing highly skilled teams of subcontractors to successfully implement complex, large-scale assessment programs. We have established successful relationships and processes with ETS, WestEd, and Measured Progress and all four of our organizations have worked together in support of PARCC. In addition, we have worked with Caveon on state assessment programs, including Kentucky and Florida with Caveon as a subcontractor for data forensics including monitoring the internet for secure content. Our previous collaborative experience will reduce the risk of launching an entirely new type of assessment program with contractors that have never previously worked together on large-scale programs:

- Pearson and ETS have a rich and successful history of working together to support state and national K–12 educational assessments. Together we currently develop and deliver assessments in California, Maryland, Mississippi, Tennessee, Texas, and Virginia, and work together to support the National Assessment of Educational Progress (NAEP). For these programs, much of ETS's work is focused on content and psychometrics, and Pearson's role is focused on the operational delivery and scoring. We are proposing similar roles and responsibilities for each organization for this PARCC project.
- Pearson and WestEd have also worked together on several assessment projects, where much of WestEd's focus has been on content development. As a subcontractor to Pearson for Phase II Item Development and a proposed subcontractor for the PARCC Diagnostic Assessments, our teams are continuing to build synergies and efficiencies around the item development process for PARCC. Additionally, WestEd has been our subcontractor for item development and forms publishing on the Arizona English

Language Learner Assessment (AZELLA) since 2011. They developed approximately 1300 items at the start of the contract and managed everything from initial test blueprints and item specifications to production of four field test forms per stage (grade band). They also managed production of CD's for the listening portion of the test. One of the other significant efforts WestEd managed was the development of a new kindergarten placement test. There, they managed all development, conducted initial cognitive labs, and managed all the production for field testing and operational materials.

- Pearson and Measured Progress currently work together in Georgia on the Georgia End of Course Test (EOCT). Measured Progress is a subcontractor providing item and forms development services for 10 specific subject areas spanning ELA, Math, Science and Social Studies. Previously, Measured Progress provided item development for us on the Georgia High School Graduation Test.
- We are currently working with Caveon and the Florida Department of Education on the statewide Florida Comprehensive Assessment Test (FCAT). Caveon is using data forensics to identify risks to security on the school and individual student level.

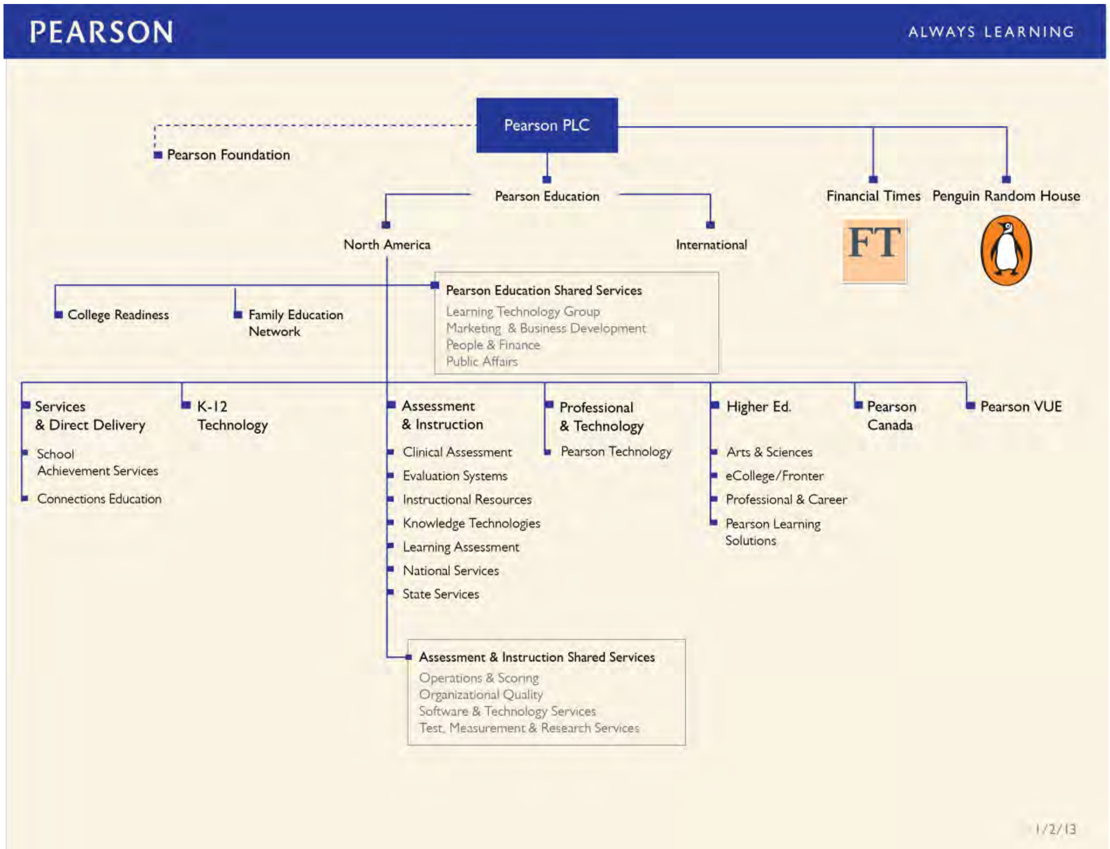
About Pearson

The Assessment & Information group of Pearson is a business unit of NCS Pearson Inc., of Bloomington, Minnesota (incorporated in 1962). NCS Pearson's address is:

NCS Pearson Inc.
5601 Green Valley Drive
Bloomington, Minnesota 55437

NCS Pearson Inc. is owned by PN Holdings Inc., which is owned by Pearson plc (a publicly held U.K. company), or its subsidiary(ies). Pearson plc (NYSE: PSO) is a publicly owned international media company with a focus on education, business information, and consumer publishing.

Based in London, England, Pearson (www.pearson.com) consists of three main groups: Pearson Education, Penguin Random House, and the Financial Times. Pearson employs more than 37,000 people in more than 60 countries around the world. The following organizational figure provides details on the business units and overall Pearson structure.



The National Services group of Pearson, a unit of Assessment & Information (A&I), will be responsible for delivery of the PARCC Operational Assessments contract. Anne Parmley, President of the National Services group, and Pat Kramer, Vice President, National Services, will oversee the project management staff assigned to this project. The project management team coordinates the work of the shared services groups that will support this project.

As part of a global reorganization, Pearson is forming a new organization called the Pearson Assessment Centre (PAC), effective January 1, 2014. This organization, to be led by Dr. Jon S. Twing, will provide comprehensive assessment services to support Pearson lines of business globally. The organization will consolidate a number of the shared service units serving PARCC, including Operations, Forms Publications, Scoring, Organizational Quality, Software & Technology Services, and Test Measurement & Research Services. This new PAC structure will promote a closer collaboration among our technology and content development resources, as well as the continued involvement of assessment resources to support the PARCC assessments.

Supporting PARCC's Needs

The Contractor for the PARCC Operational Assessment contract needs to:

- Develop a large number of passages and items for ELA/literacy (grades 3-11) and mathematics (grades 3-8 and end-of-course and integrated mathematics high school courses)
- Support states in the transition to online testing using the PARCC Assessment System
- Provide paper-based and online tests, manuals, and practice tests
- Develop operational and embedded field test forms production
- Provide services for registration, materials printing, pre-ID, packaging, distribution, and processing
- Provide performance scoring and artificial intelligence scoring services
- Conduct psychometric analysis, research and standard setting
- Provide operational assessment score report services
- Provide call center and technical support services

Outstanding Item Development Team

An experienced Pearson, ETS, and WestEd item development team offers know-how and the capacity to deliver grade-appropriate items of sufficient quality to survive a rigorous internal and, if required, external item review—which means more efficient item development. PARCC can rely on the vast experience of this collaborative to deliver items for assessments that are designed to be fair, valid, and reliable. Pearson alone supports large-scale assessment development projects for more than 20 states, two national testing organizations, and one federal agency. We develop approximately 40,000 test items a year. Pearson's experience is rooted in a combination of qualified content and assessment development professionals, solid item development processes, and a strong team-oriented approach to developing and delivering items which will be powerful tools in the teachers' classrooms.

Best Practices for Item Development

Well-written items and forms result not only from proven processes and talented staff, but from adherence to recognized guidelines for assessment development. Pearson follows the *Standards for Educational and Psychological Testing* (American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME),1999) and industry best practices in developing selected-response and constructed-response items, as well as performance tasks.

Multiple reviews built into our item development process also lead to quality items that will aid the teachers in determining whether the students have a firm understanding of the unit and objectives.

> *The collaborative of Pearson, ETS, and WestEd's content specialists have formal training in their areas of expertise as well as extensive classroom and large-scale assessment experience.*

Practical experience contributes to creating items that perform to PARCC needs and expectations. Members of this collaborative content and assessment specialist team not only have formal training in their areas of expertise, but have extensive classroom and large-scale assessment experience. Additionally, staff receive training in designing clear, precise, non-biased items that are amenable to accommodations and accessible to the breadth of a given student population. This background and training enables these content and assessment specialists to produce high-quality items that are favorably received by customers and have a high approval rate through the customer review process.

Item Development for Multiple Assessment Modes

High-stakes and classroom assessments are undergoing a major transition. Many states are moving to online testing, not just for their high-stakes programs but also for their classroom assessments which they may administer through an instructional improvement system.

When assessments are administered online, this opens up a range of possibilities with respect to item layout and student engagement. Technology enhanced "innovative" items and/or performance tasks, can provide a way to take an activity and bring it to life on the computer screen in a way that can go beyond the traditional paper and pencil tests.

To arrive at methods suited for online assessments and innovative item development, we have modified our traditional processes for developing paper-based tests. For instance, to reduce the costs associated with programming complex simulations and detailed animations, Pearson has implemented a series of customer reviews to validate and refine the design before we develop fully functioning items.

Our emphasis on template development also has reduced development costs and the expenses associated with quality control testing. Releasing sample items based on our templates has enabled students to become familiar with the interface associated with various item types before taking the test. By using these pre-existing templates for item development, we can also greatly reduce the time and the cost of these items.

If PARCC desires, we can discuss the process and costs of developing innovative items for an online instructional improvement system.

Expert Test Development and Analysis

At Pearson, all test development takes place within our Test, Measurement, and Research Services (TMRS) group rather than across several departments. Having our content experts and psychometricians in the same department facilitates communication, project planning, and close collaboration for the most efficient approach to test development. Our psychometric and content development teams already know how to work with each other to PARCC's benefit. Our subcontractors ETS and WestEd have similar test development processes and we will work as a team to standardize these processes so the final look and feel of the items is consistent for PARCC across all grades and subjects.

At Pearson, our content development and psychometric research capabilities come together to provide the following:

Pearson Content Development and Psychometric Research Capabilities	
Test blueprints	Pilot, field test, and operational test design, sampling, and data analysis
Statistical analyses	Validity and reliability investigations
Item development aligned to state standards	Differential Item Analysis (DIF) and other statistical data analyses, such as item and scaling drift
Item reviews for content, bias, and universal design	Standard setting
Test form development, including forms adapted for online administration	Accommodated testing of English as a second language populations and examinees with disabilities
Automated item banking, tracking, and test forms construction	Preparation of technical documentation and reports
Calibration, scaling, and equating using a broad array of classical and IRT-based measurement models	Presentations of technical data in a variety of media and to various audiences, including state boards of education and state legislators
Special studies to examine various assessment issues (e.g., comparability studies, generalizability analyses, factor analytic studies, cognitive labs)	Support for states as they defend and explain how they are meeting requirements of Federal Peer Review and measurement best practices

Developing State Assessments. Pearson psychometricians and content development specialists support our state customers through use of sound methodologies in all areas of test development.

Our established processes allow our staff to develop items or work with educators to develop items that align to applicable content standards, adhere to the principles of Universal Design, and reflect best instructional practices.

Dr. Jon S. Twing, Pearson's Chief Measurement Officer, directs four functional teams:

- Business Management, led by Karen Squires Foelsch
- Measurement Services, led by Dr. Walter "Denny" Way
- Research and Innovation Network, led by Kimberly O'Malley
- Learning Integration, led by Dean Brown

Measurement Services

The Common Core State Standards provide guidance about what students should understand and be able to do at critical points in their education.Developing tests that fairly and reliably measure student learning for these new standards takes experience and capacity—hallmarks of Pearson capabilities and are further augmented by our subcontractors ETS and Measured Progress. To support item and test development and test administration in a changing educational landscape, Pearson provides customers with the content and psychometric support shown in the following figure.

Measurement Services for Pearson State Customers	
Initial evaluation of test items and prompts, including alignments	Accommodated versions of an assessment
Content development in all major subject areas, including general population, English language learners, and special needs students	Quality control activities to maintain consistency across all item/test form development, administration, scoring, and reporting functions
Incorporation of stimuli, items, and prompts into pilot and field test forms	Development, maintenance, and support for test security measures
Participation in internal and external item/prompt content, bias, and data review	Securing of permissions, payment on behalf of clients, tracking, and records handoff
Item/prompt scoring and subsequent analyses of pilot test data	Descriptions of test administration procedures
Test form development, including descriptions of test blueprints and alignment to clients' learning standards, practice forms, and descriptions of their contents	Training and other consultation

Supporting Assessment Quality. Our established processes enable our staff to develop items that align to client content standards, adhere to the principles of Universal Design, and reflect best instructional practices.

To give each student the best possible opportunity to demonstrate his or her full range of knowledge and skills requires a full range of Pearson personnel—including artists, content specialists, researchers, copyeditors, and fact checkers. Many of our staff hold teaching certifications and have experience teaching in K–12 classrooms. Their commitment to quality education shows in the care they give to creation, review, and production of every test item.

This care results also from established processes developed over decades of assessment experience. Our procedures enable Pearson content specialists to develop items that align to client content standards, adhere to the principles of Universal Design, and reflect best instructional practices, including the 1999 AERA, APA, and NCME Standards for Educational and Psychological Testing. Applicable APA Standards guide test development.

Industry-Leading Print Services

Our experience in scannable forms design and printing will deliver quality and reliability to PARCC. Because precise printing standards for scannable forms on registration, image size, and trimming must be rigorously upheld for forms to scan properly, we have developed this capacity in-house. Our careful adherence to higher standards produces forms that scan accurately, keeping PARCC assessments on schedule.

> *PARCC can rely on Pearson for superior printed materials. Our print services are certified to ISO 9001:2000 standards.*

As a demonstration of our adherence to quality standards in printing, Pearson print servicesearned the internationally recognized ISO 9001:2000 certification after an independent audit by the ISO Registrar. At the time we first earned ISO-certification in 1994, it was unprecedented in the printing industry. Our attention to production detail throughout all phases of printing produces documents that deliver consistently reliable scans.

Because we print all our scannable forms internally, we can offer PARCC flexibility in document design, while still using our proven techniques. Additionally, printing internally allows us to deliver a professional product at a lower cost. Our facilities annually undergo multiple external and internal audits to verify process compliance.

Along with quality, Pearson also offers the press capacity critical to short testing windows. Our printing plants are able to produce more than 1 billion sheets per year. We remain the nation’s largest supplier of scannable forms for high-stakes educational testing. We design forms to optimize data collection and processing, depending on how our customers will use their data.

Efficient, Accurate Test Delivery

Pearson provides orderly, accurate, and timely packaging and distribution. We use barcodes, so we know where all secure documents are and where they should be going. We track shipments to verify they have been delivered to the right location so your students receive the right tests, at the right time. When materials arrive at testing sites, test coordinators receive organized shipments, making it easier for them to distribute tests to the appropriate grades.

We assign barcodes at multiple stages of packaging for efficient materials tracking. Before packaging, we print a unique barcode serial number on secure test booklets and on other materials that require security.

We use an inline quality control system to verify that barcodes are printed sequentially on materials. We group and weigh materials prior to wrapping to verify they are correct.

We take non-secure materials that were not barcoded and assemble them in specific package types and sizes to prepare them for final packaging. We then track these materials by a barcode on an assembly identification sheet, which we use to help with assembly during pre-packaging. This sheet identifies all the items within that assembly.

Our pick and pack process provides accurate pallets and packing slips. During final packaging, Pearson uses scanners with a system-generated pick slip to select the materials required for each school. Our system generates a packing slip and pallet detail report for each shipment.

To verify that the correct materials go to the right destination, scanners we use during our pick and pack process prompt packing personnel to scan the requested item in the correct packing order. The scanner alerts personnel if they have scanned the wrong barcode.

We meet schedules with on-time shipping. With our packaging and distribution system, we aim for complete shipment accountability from origin to destination to return. Although we have established successful relationships with many shipping vendors, we work primarily with UPS, which provides specialized handling and delivery services that help us maintain test security and meet your schedule.

We have the capacity to meet your shipping needs. In 2012, we scanned and shipped more than a million packages of assessment materials, each with a unique barcode for tracking.

We provide for straightforward test materials check-in. To help testing administrators easily check shipments for accuracy when they arrive, we include easy-to-read shipping reports. Packing lists and pallet detail reports provide accurate information about the quantities of materials we boxed and to what locations we shipped them.

Fast and Secure Test Processing

Large-volume processing capabilities, advanced technologies, and experienced personnel enable Pearson each year to process, score, and report immense volumes of documents for student assessment programs in an accurate and timely manner.

We scan approximately 8 million pages per day. Documents containing both student demographic data and test item responses are scanned, edited, scored, and reported annually for assessment programs in multiple states. This means we have the capacity to scan, edit, score, and report for the PARCC Operational Assessments quickly and accurately. The process of scoring machine-scannable documents begins when school districts across

PARCC states ship their scorable materials to Pearson. It ends when they receive their assessment report results.

Secure, High Capacity Facilities

Pearson activities in support of large-scale K-12 assessments primarily occur in our Iowa, Minnesota, and Texas facilities. Together, they provide more than 1.3 million square feet where we develop, design, print, package, process, scan, score, warehouse test materials, report results, and provide administrative support.

Our technology and quality processes keep Pearson program management teams informed as work is performed at multiple facilities.

Pearson Facilities Supporting Large-Scale Assessment

Location	Types of Facilities	Facility Processes	Facility Size / Operating Hours	Key Features
Boulder, CO 	Offices	▪ Automated Scoring Development ▪ Research and Development ▪ Project management	18,000 square feet	Research and development
Iowa City, IA 	Offices Data center Processing center Performance scoring center	▪ Program management ▪ Item/ test development ▪ Psychometrics ▪ Pre-press ▪ Software development ▪ Scanning/ data editing ▪ Machine/ human scoring	339,000 square feet Operates 24 x 7	Data Center housed in disaster-proof structure has massive capacity
Cedar Rapids, IA 	Distribution center Warehouse Printing Operations	▪ Packaging/ distribution ▪ Non-scannable printing ▪ Pre-mailing ▪ Warehousing/ material storage ▪ Reports printing ▪ Reports assembly and delivery	265,000 square feet Operates 24 x 7	Transformable work space allows us to meet changing demands with ease
Owatonna, MN 	Printing plant	▪ Scannable/ non-scannable printing ▪ Offset/ digital printing ▪ Distribution	128,000 square feet Operates 24 hours x 5 days per week	Capacity to produce more than 1 billion scannable sheets each year

Pearson Facilities Supporting Large-Scale Assessment				
Location	**Types of Facilities**	**Facility Processes**	**Facility Size / Operating Hours**	**Key Features**
Austin, TX 	Performance scoring center Offices Processing center Warehouse Distribution center	▪ Program management ▪ Item/ test development ▪ Psychometrics ▪ Packaging/ distribution ▪ Scanning/ data editing ▪ Data processing ▪ Machine/ human scoring ▪ Warehousing/ material storage	280,000 square feet Operates 24 x 7	Housed in a technology corridor where we research innovative solutions to better serve our customers
San Antonio, TX 	Performance scoring center Offices Warehouse Distribution center	▪ Program management ▪ Test / item development ▪ Psychometrics ▪ Pre-press ▪ Human scoring ▪ Warehousing/ material storage	575,000 square feet Operates 24 x 7	Large meeting rooms for range finding and sensitivity meetings, keeping costs low

Pearson Facilities Supporting Activities for Large-Scale Assessment. Pearson maintains highly secure facilities, with access restricted to authorized personnel. Visitors must be escorted at all times. Multiple processing centers afford us additional backup capacity if the need arises.

ISO-Certified Performance Scoring

Our scoring capabilities provide a full spectrum of subject areas for constructed-response scoring. In addition to scoring constructed-response items, we also offer rangefinding/benchmarking services to set appropriate scoring standards; professional development workshops for educators; evaluation of item prompts from field tests for scorability; and scoring in other media, such as audio- and video-taped responses.

Pearson offers the capacity and expertise necessary for large-scale assessments. Our scoring contracts include more than 18 state programs, as well as Washington D.C. and Puerto Rico programs, the National Assessment of Educational Progress (NAEP), and National Board for Professional Teachers (NBPTS) assessments.

More Innovation: Pearson Performance Scoring

PARCC will benefit from our capacity and proven ability to adapt, extend, and apply technologies in new ways that advance educational assessment. Pearson was one of the first organizations to use an image-based scoring system to support human scoring. We went on to develop, refine, and expand this technology to reduce paper handling, promote reliability and validity testing, and improve scoring turnaround time.

Pearson was the first in scoring with these many innovations:

- **FIRST** with the capability to present scoring organized by prompt
- **FIRST** to electronically track scorer performance to provide feedback (on speed, inter-rater reliability, training, and so forth)
- **FIRST** to develop a dynamic online scoring guide so scoring rules display with the item to be scored
- **FIRST** to institute a system for second and third resolution readings, improving reliability
- **FIRST** to institute transparent scoring of calibration responses to measure scorer drift, increase inter-rater reliability, and improve validity
- **FIRST** to be ISO certified for performance scoring

Ample Scoring Capacity

Our proprietary image-based scoring system enables us to score and monitor projects of all sizes, from small to large and complex. With multiple scoring support sites in addition to our proprietary distributed scoring system, we have scoring capacity for projects large and small. We also score specialized projects, such as small paper-based item and task tryout pilots or field tests using a data capture system to support analyses and review by our content developers and our psychometrics team.

User-Friendly Reports

Pearson reports aim to enhance learning at the student, classroom, and district levels. It is not enough to just generate reports. They must be comprehensible, and reflect student performance in clear and concise fashion. We will work with PARCC and the data warehousing and reporting vendor to not only generate reports but enhance their usefulness.

Stringent Security for All Activities

Proper security is essential to protect the integrity of our customers' data, from beginning to end of a program and at all steps in between. To reduce our customers' security concerns, we have implemented multiple security features, as shown in the following figure.

Pearson Assessment Security	
Facilities	▪ Badge access ▪ Supervised visitors ▪ Guards on duty 24 hours a day, 7 days a week ▪ Closed-circuit cameras
Test Development	▪ Signed security agreements to protect confidentiality of all items and forms ▪ Proprietary item development and test management system with stringently controlled access ▪ Secure control of and accountability for all meeting materials
Data and Electronic Transfer	▪ Secure file transfer protocol sites ▪ Offsite data storage and backup ▪ Strict password rules for access
Subcontractors and Vendors	▪ Stringent security protocols with audits ▪ Non-disclosure agreements ▪ Secure storage of materials
Test Administration	▪ Online test lockdown to prevent cheating ▪ Marked boxes for secure materials with detailed instructions for district test coordinators ▪ Security agreements and surveys for test administrators
Distribution and Materials Return	▪ Barcodes for tracking of and accounting for all secure materials ▪ Security-certified carriers

Stringent Security. Providing security is critical to maintaining the technical quality, perceived fairness, and integrity of an assessment program.

Rigorous Quality Assurance Processes

Our clients depend on our integrity and accuracy. Our quality control focuses on defining and implementing critical processes so we can deliver products and services to our clients that meet or exceed their requirements.

Proven methodology helps us deliver on our promises. For example, the PMI project management model provides processes for quality checks throughout a project's life cycle, from planning and development through implementation and renewal or close-out. We track performance metrics for productivity and quality in our production areas. Each month, we collect, present, and discuss these metrics at a management team review. Production department managers, process engineers, and members of our quality team use data to identify the causes of errors, barriers to quality and productivity, and areas for process and quality improvement.

Using PMI methodology, we have established repeatable processes for project team coordination, formal training on a common process, shortened learning curves, and complete documentation of program processes. Adhering to the PMI model produces predictable and consistent results.

PMI Program Management

Perhaps the most visible employees to our customers are the members of our program teams. Pearson program teams follow the guidelines of the Project Management Institute (PMI®), an international body of knowledge that provides a consistent management framework. Our program managers also possess or are actively pursuing PMI Project Management Professional (PMP) certification.

The PMP certification program objectively assesses and measures professional knowledge. To achieve PMP certification, candidates must satisfy PMI's educational and professional experience requirements and demonstrate a proper level of understanding and knowledge of project management, which is assessed by the Project Management Professional Certification Examination.

In addition, those who have earned the PMP credential must demonstrate ongoing professional commitment to the field of project management by satisfying PMI's Continuing Certification Requirements Program.

Our program managers bring a variety of experience to their positions. Some have classroom or education administration experience, while others have worked for years in different areas of Pearson or other testing companies. All of this experience brings critical education and testing knowledge to the programs they manage.

PMI Methodology Will Benefit PARCC

PARCC will benefit from our adherence to the PMI management model, which promotes quality and on-time delivery throughout a project's duration. The model's framework encourages our program team to determine program requirements, adhere to a detailed project plan, effectively manage scope changes, and follow sound business practices to minimize risk. In turn, this helps us achieve greater predictability and repeatability of results.

Efficient, Responsive Customer Service

Our Customer Service Center (CSC) will promptly respond to the questions from PARCC states. Our CSC uses call routing and a team approach which together allow us to deliver excellent, efficient service across numerous assessment programs with varied administration windows.

When PARCC test coordinators call for customer support, they will want to talk to a live person who is a qualified professional and able to offer immediate assistance. Pearson provides experienced, competent customer support staff who can answer questions efficiently and thoroughly. We require that our CSC staff demonstrate excellent customer service skills in addition to appreciating that each caller is an individual with specific needs and concerns.

Pearson References

References from Maryland, Minnesota, and Virginia are provided in the following figures. Appendix H, Reference Questionnaire is being completed by each referenced customer. The completed forms will be returned directly to the New Mexico Procurement Manager, as required in the RFP.

State	Maryland
Contract	**Maryland School Assessment (MSA)** 2002–2014 reading and math, 2006–2014 science **Alternate Maryland School Assessment (Alt-MSA)** 2003–2015 **Maryland High School Assessment (HSA)** 2007–2013 (sub) **Maryland Functional Testing Program (MFTP)** 2002–2008
Overview	**MSA:** Students test in reading and math at grades 3–8 and in science at grades 5 and 8. From 2009–2012, modified reading and math tests **(Mod-MSA)** were given to students unable to participate in the MSA, even with accommodations. **Alt-MSA:** Students with significant cognitive disabilities are tested in reading, math, and science at grades 3–8 and 10. **HSA:** Students take end-of-course exams in English, government, biology, and algebra/data analysis. Modified tests **(Mod-HSA)** are given to students unable to participate in the HSA, even with accommodations. **MFTP:** Students with grade 12 status as of the 2003–2004 school year were required to pass online computer-adaptive tests in reading and math to graduate.
Products and Services Provided by Pearson	▪ Comprehensive program management ▪ Item/form development ▪ Psychometric services, including analysis, standard setting, and technical reports ▪ Training meetings ▪ Online professional development for teachers (Alt-MSA) ▪ Online mastery objective bank (Alt-MSA) ▪ Review of mastery objective targets written by teachers (Alt-MSA) ▪ Test materials printing and distribution ▪ Online testing ▪ Administration manuals ▪ Scanning, scoring, and performance scoring ▪ Automated scoring of constructed-response items (MSA science) ▪ Scoring student portfolios (Alt-MSA) ▪ Results reporting
Client Contact	Janet Bagsby Section Chief, Planning and Assessment Branches Division of Accountability and Assessment Maryland State Department of Education 200 W. Baltimore St. Baltimore, MD 21201-2593 410.767.0048 jbagsby@msde.state.md.us

State	Minnesota
Contract	**Minnesota Comprehensive Assessments (MCAs)** 1998–2011 **Minnesota Basic Skills Tests (BSTs)/Graduation-Required Assessments for Diploma (GRAD)** 1999–2010 **Minnesota Test of Emerging Academic English (TEAE) and Mathematics Test for English Language Learners (MTELL)** 2005–2010
Overview	**MCAs:** Students are tested in reading and math at grades 3–8, 10, and 11. Science tests are administered at grades 5 and 8 and the year in high school when students complete a life science course. Students are tested on writing skills in grade 10. Reading tests are paper-based, science tests are computer-delivered, and math testing is administered in both formats. **BSTs/GRAD:** To be eligible for a high school diploma, students who entered grade 8 in 2004–05 or earlier were required to obtain passing scores on the BSTs in reading, math, and writing. Students entering grade 8 in 2005–06 or later must pass the GRAD in writing at grade 9, reading at grade 10, and math at grade 11. **TEAE and MTELL:** English language learners are assessed at grades 3–8 and 11 in reading, writing, and math.
Products and Services Provided by Pearson	▪ Project management ▪ Psychometric consulting and conducting standard setting meetings ▪ Item development ▪ Test form equating ▪ Conducting rangefinding activities with state staff and selected Minnesota educators ▪ Review of scoring rubrics ▪ Field testing of matrix items to build future forms. ▪ Design, printing, and distribution of test booklets, answer documents, and manuals ▪ Online testing ▪ Performance scoring for writing responses and constructed-response items in reading and mathematics ▪ Results reporting at the student, school, district and state levels
Client Contact	Jennifer Dugan, Director Minnesota Department of Education 1500 Highway 36 West Roseville, MN 55113 651-582-8654 Jennifer.Dugan@state.mn.us

Customer	Virginia
Contract	**Virginia Standards of Learning Program (VASOL):** 2001–2014 **Virginia Grade Level Alternative Assessment Program (VGLA):** 2006–2014 **Virginia Substitute Evaluation Program (VSEP):** 2005–2014 **Virginia Alternate Assessment Program (VAAP):** 2005–2014 **Virginia Modified Achievement Standards Test (VMAST):** 2005-2014 **Virginia Algebra Readiness Diagnostic Test (ARDT):** 2012-2014 **Virginia Educational Information Management System (EIMS):** 2001–2014

Customer	Virginia
Overview	**VASOL:** Science and history at grades 3, 5, 8, and high school. Students take these Students are assessed in English and math at grades 3–8 and high school and in tests primarily online though paper/pencil tests are available for students with a documented need. **VGLA:** Available for certain Limited-English Proficiency students and students with disabilities in grades 3-8 as an alternative assessment for the SOL testing. Students can participate in this assessment for, science, history, and writing at grades 3–8. (Based on collection of evidence.) **VSEP:** An alternative method of assessing students who by nature of their disability are unable to participate in the end-of-course SOL assessments even with testing accommodations. (Based on collection of evidence.) **VAAP:** Designed to evaluate the performance of students with significant cognitive disabilities who are working on academic standards that have been reduced in complexity and depth. Students are tested in reading, math, science, and history at grades 3–12. (Based on collection of evidence.) **VMAST:** Students with disabilities who are being instructed in grade level content but are not likely to achieve proficiency in the same time frame as their non-disabled peers at grades 3-8 reading and math as well as end-of-course reading and algebra I. Solely an online test. **ARDT:** Provides mathematics intervention services to students in grades 6-9 who are at risk of failing the Algebra I end-of-course test. **EIMS:** Pearson developed, implemented, and hosted the user friendly data platform to manage Virginia's education information statewide. The Virginia EIMS uses a unique State Testing Identifier (STI) to follow assessments given to each of the 1.3 million students in the Virginia school system.
Products and Services Provided by Pearson	▪ Item/form development ▪ Psychometric services including standard setting, scaling and equating, and running student growth percentile data ▪ Online and face-to-face training for districts and schools ▪ Assigned a unique state testing identifier to more than 1.2 million students through the Virginia Educational Information Management System (EIMS) ▪ Test materials printing and distribution ▪ Web-based "Understand Scoring" module (teachers learn how professional scorers apply the statewide rubric to student writing responses) ▪ Teacher training for scoring collections of evidence (VGLA and VAAP) ▪ Single sign-on portal for online test management ▪ Online testing and online scoring ▪ Performance scoring (distributed and regional) ▪ Data analysis ▪ Results reporting—aggregated, disaggregated, comparative, longitudinal ▪ E-transcripts for higher education institutions ▪ Information management systems

Customer	Virginia
Client Contact	Shelley Loving-Ryder Assistant Superintendent for Assessment and Reporting Virginia Department of Education 101 N. 14th St. James Monroe Bldg., 18th Floor Richmond, VA 23218-2120 804.225.2102 Shelley.loving-ryder@doe.virginia.gov Dr. Lan Neugent Assistant Superintendent for Technology 804.225.2757 lan.neugent@doe.virginia.gov

About Caveon Test Security

Qualifications and Expertise

In 2003, a group of test industry veterans founded Caveon for one sole purpose: To help protect the items, programs, and reputations of their clients—including large, high stakes testing programs—by enhancing test security.

Over the years, Caveon has worked closely with a large number of states and districts to assist them with comprehensive and thorough audits of their testing programs, analyses of test data to achieve trustworthy test results, and web monitoring of the internet to detect test security threats.

Caveon has also been involved in several high profile investigations of testing irregularities and possible cheating, in particular, the work done in 2010–11 for the State of Georgia and the Atlanta Public Schools and the work done in both 2009–10 and 2010–11 for the District of Colombia Public Schools. Caveon's work on these two programs was crucial to help uncover cases of unethical behavior. In addition, Caveon has conducted investigations on behalf of the Colorado Department of Education.

Caveon fully understands that it is crucial to maintain a high degree of ethics in the state educational system and that all teachers must be held to an honor code. The implementation of a Caveon Web Patrol™ web monitoring program for the PARCC consortium will greatly enhance the security of the assessments and the integrity of test scores used for accountability purposes.

Because of all the work Caveon has done with states and districts in recent years, they are very aware of all the issues related to test security and their connection to state/district procedures, policies, and the materials used with testing programs. Caveon staff and consultants have much experience and expertise in working with education assessment programs on a variety of issues, as well as with major testing vendors and other organizations involved in high stakes testing. Some of our accomplishments include the following:

- Caveon staff are often asked to provide advice on how to prevent cheating, and Dr. John Fremer has been interviewed many times by reporters for national news sources (*USA Today*, National Public Radio, *The New York Times*, etc.) to comment on this topic.
- In September of 2011, key Caveon personnel (Dennis Maynes, Chief Scientist) participated in a special meeting with high-level officials from the US Department of Education to provide input on the topic of increased cheating on state NCLB assessments. The US Department of Education recognized the value and expertise that Caveon provides in this area and invited staff to share critical information on analysis and detection procedures.
- Dr. David Foster, Caveon CEO and Chairman of the Board, has led the re-write effort on the Security Chapter for the Council of Chief State School Officers (CCSSO) "Operational Best Practices for Statewide Large Scale Assessment Systems."
- Caveon's Dr. Fremer and Dr. John Olson co-authored the recently published CCSSO State Collaborative on Assessment and Student Standards (SCASS) "TILSA Guidebook on Test Security," which focuses specifically on implementing data forensic programs to address test security issues.
- Both Drs. Foster and Fremer participated as expert panelists in US Department of Education's National Center for Education Statistics' 2012 Testing Integrity Symposium in Washington DC.
- Dr. John Fremer co-edited the recently published "*Handbook of Test Security*."

For state assessment programs, Caveon Security Audits, Caveon Data Forensics, Caveon Web Patrol, and/or Caveon Security Investigations have been conducted with the following state departments of education and large school districts:

- Colorado DOE
- Delaware DOE
- Florida DOE
- Idaho DOE
- Indiana DOE
- Kansas DOE
- Kentucky DOE
- Louisiana DOE
- Massachusetts DOE
- Minnesota DOE
- Oregon DOE
- Pennsylvania DOE
- South Carolina DOE
- Texas DOE
- Washington, State OSPI
- Wisconsin DPI
- Atlanta Public Schools
- Baltimore Public Schools
- Dallas Intermediate School District
- Durham (North Carolina) Public Schools

- Mississippi DOE
- Nebraska DOE
- North Carolina DPI
- Fairfax County Public Schools
- Washington, DC Public Schools

References for Caveon Test Security

State	Florida Department of Education
Contract	**Florida Comprehensive Assessment Test (FCAT)**
Overview	FL DOE is a current client, utilizing Caveon Data Forensics to identify risks to security at both the school and individual student level for the statewide Florida Comprehensive Assessment Test (FCAT) program. Our service is part of DOE's quality assurance process for scoring. In addition, Caveon conducted two security audits (one for state assessment, the other for the FL Teacher Certification program) in 2006.
Client Contact	Victoria Ash Bureau Chief K12 Student Assessment Victoria.Ash@fldoe.org Phone: 850-245-5513

State	Mississippi Department of Education
Contracts	Subject Area Testing Program (SATP), Mississippi Curriculum Test, Second Edition (MCT2)
Overview	Since 2006, MDE has utilized Caveon Data Forensics statewide for identifying security risks to administrations of its high-stakes Subject Area Testing Program (SATP). This year, the engagement has expanded to also include its Mississippi Curriculum Test, Second Edition (MCT2). In addition, in 2013, Caveon will implement data forensics for the Mississippi Science Tests (MST2) and Subject Area Test Program (SATP2).
Client Contact	Walt Drane Assistant Director, Office of Student Assessment WDrane@mde.k12.ms.us Phone: (601) 359-3052

State	Kentucky Department of Education
Contracts	**English Language Arts and Mathematics for grades 3-8**
Overview	In the last year, Caveon has conducted both a Security Audit and a pilot Data Forensics analysis to ensure trustworthy test results for English/Language Arts and Mathematics for grades 3-8.
Client Contact	Roger Ervin System Director, IT Office of Assessment and Accountability roger.ervin@education.ky.gov Phone: 502-564-2256 ext. 4719

State	North Carolina Department of Public Instruction
Overview	Caveon has conducted both a Security Audit and Data Forensics for NC DPI.
Client Contact	Lou Fabrizio Director of Accountability Policy and Communications lfabrizio@dpi.state.nc.us Phone: (919) 807-3770

About ETS

ETS Corporate Capabilities

ETS is a nonprofit corporation started in 1947, whose mission is to advance quality and equity in education for people worldwide by creating assessments based on rigorous research.

As one of the world's largest private educational testing and measurement organizations, ETS develops, administers, or scores more than 50 million tests annually in nearly 200 countries at more than 25,000 locations. Because of ETS's experience and the quality of their staff, ETS is qualified to design, develop, and implement customized assessment systems and support services tailored to meet the needs of clients and students.

Commitment to PARCC

ETS is committed to PARCC and its success. As one of PARCC's two item development prime contractors, and as one of the contractors supporting PARCC's assessment administration, ETS is well-versed in the PARCC project and the needs of the client. The work ETS has completed with Pearson on current PARCC contracts has allowed our team to demonstrate exceptional content and technical expertise, our understanding of and ability to support the complex and dynamic needs of the client, and an ability to work in close collaboration in support of PARCC.

ETS fully supports Pearson on the operations related to delivering the PARCC Operational Assessment Program and is committed to delivering the assigned scope of work and accepting the associated obligations as detailed in this proposal.

Summary of Related Experience and Qualifications

Through the Student and Teacher Assessment division, ETS has developed relationships with state departments of education, their administrators and staff, as well as the educators within the states' local education agencies. The existing K–12 relationships also include organizations that represent multiple states and geographic regions. The PARCC and Smarter Balanced consortia have created important new opportunities for partnership and innovation across the assessment industry.

ETS's K–12 roots are in customized large-scale, high-stakes state assessments. It has 28 active contracts in its K–12 portfolio, the majority of them with individual states. ETS serves as both the prime contractor, where it manages and coordinates the activity of other vendors, or as a subcontractor to other companies.

For the past 11 years ETS has been the prime contractor for the Maryland High School Assessments. The Maryland State Department of Education has acknowledged ETS's high level of service by awarding it multiple contract extensions, and giving ETS a rating of 10 out of 10 on recent yearly client satisfaction surveys. ETS has also been the prime contractor for the California High School Exit Examination (CAHSEE) since 2001, and the California Standardized Testing and Reporting (STAR) program since 2002. Combined, these California programs test more than 6 million students annually. California and Maryland represent the kind of collaborative long-term relationship that enables ETS to be of the greatest service to its clients.

Assessment Development teams at ETS have worked as a subcontractor to Pearson for more than 14 years. Current Assessment Development contracts include: the Virginia Standards of Learning (SOL) testing program; the Tennessee Comprehensive Testing Program (TCAP), where we develop both the 3-8 Achievement and high school End of Course exams; the Mississippi Subject Area Testing Program 2 (SATP2); and the State of Texas Assessment of Academic Readiness (STAAR).

ETS develops its scope of work in a variety of ways, according to the needs of its customers. For Virginia alone, in previous years ETS has developed more than 10,000 items and 800 test forms annually for 34 different subject area tests. A list of ETS's current contracts is at the end of this section. Client references highlighting ETS's relevant scope for three contracts are provided along with client contact information.

ETS Professional Staff

ETS employs more than 3,000 people, including 1,100 professional staff with training and expertise in education, instrument development, psychology, statistics, and psychometrics. Additional areas of staff expertise include the computer sciences, sociology, and the humanities. These professional staff will support the areas that directly impact ETS's scope of work for this proposal:

- Assessment Development
- Data Analyses, and Psychometric Analyses and Research
- Validity Research: Accessibility
- Program Management

The following descriptions provide an overview of the three core divisions that support the PARCC Operational Administration scope of work.

Assessment Development

With more than 600 full-time staff, many of whom are former teachers, ETS Assessment Development is responsible for developing some 200 test titles for ETS clients. Its Assessment Development division includes five areas:

- K–12 Assessments
- General Skills Assessments
- Specific Subject Assessments
- National Assessment of Educational Progress
- English Language Learning

ETS develops, analyzes, and validates content according to each testing program's specifications and according to guidelines that it bases on recognized standards in the field of educational measurement.

Housed within the Research and Development Division and led by Dr. Marisa Farnum, the ETS Assessment Development Division is responsible for the development of items, tests, and related materials for all of ETS's major brands and clients.

In close cooperation with clients, ETS provides the following services related to test development:

- Design
  - using Evidence-Centered Design, develops content and domain frameworks to underpin assessment blueprints and specifications
  - designs test blueprints, test specifications, item specifications, and scoring materials
  - designs and develops new, innovative item types based on measurement theory and cognitive science in response to client and industry-wide needs
- Item and Test Creation and related training
  - creates high-quality item content using highly trained subject matter experts
  - conducts teacher item writing workshops and external item writers
  - transforms draft items into test-ready items, using multistage processes including content, editorial, bias, and accessibility quality-control reviews
  - conducts content, bias, and sensitivity reviews with educator committees for newly written items
  - conducts statistics reviews with educator committees for piloted items
  - assembles test forms, including directions for test takers and test administrators
  - produces computer-delivered and paper-based test forms
  - develops practice tests and sample items
- Post-assessment activities
  - reviews test results with the ETS Statistical Analysis team to identify potential bias issues or testing anomalies
  - prepares technical reports
  - publishes tests and items for public release
  - exports tests and test items using a variety of data formats, including QTI 2.1 and APIP

ETS also has created and delivered other innovative item formats, including online scenario-based tasks, portfolio tasks, speaking-proficiency tasks, and writing-proficiency tasks, as well as a host of technology-enhanced item types. ETS remains on the cutting edge of test item design and experimentation to meet emerging needs, and it supports all the primary technology-enhanced item behaviors defined in the QTI 2.1 standards. Currently, ETS is working with formative assessment scenarios and questions that offer learning opportunities to students. ETS has been conducting a long-term research and development initiative called Cognitively Based Assessment *of*, *for*, and *as* Learning (CBAL™) that aims to create a future comprehensive system of assessment that documents what students have achieved (“*of* learning"), helps identify how to plan and adjust instruction ("*for* learning"), and is considered by students and teachers to be a worthwhile educational experience in and of itself ("*as* learning").

ETS reviews its test materials according to the guidelines in the ETS Fairness Review Guidelines, which call for test developers to design and produce content that is likely to give all test takers an equal opportunity to fairly demonstrate their knowledge, skills, and abilities. ETS also has published guidelines for international assessments, the testing of K–12 English language learners, and the development and scoring of performance tests. Please see www.ets.org/fairness for more information about the organization's standards and guidelines.

Data Analysis, Psychometric Analysis, and Psychometric Research

ETS devotes more resources to research than any other education assessment company, and it will share the relevant research with Pearson and PARCC to support the program's success. ETS has developed educational measurement research and analysis, innovative product development, and original policy studies.

The ETS research group is staffed by more than 180 professionals, including some of the nation's most distinguished scientists from the fields of psychometrics and statistics. Many of the now-common psychometric procedures — such as the Angoff series of equating models, item response theory (IRT), and differential item functioning (DIF) — were pioneered at ETS.

The ***Statistical Analysis, Data Analysis, & Psychometric Research*** group, led by Dr. John Mazzeo, provides both standard psychometric analyses to support ongoing programs and development/deployment of cutting-edge psychometric processes to support innovative assessments.

ETS research staff provides psychometric support for assessments and measurement-related services that are standards of excellence in the testing industry. ETS supports the technical standards described in the AERA/APA/NCME Standards for Educational and Psychological Testing and creates its own stringent set of policies to which its entire organization must adhere: the ETS Standards for Quality and Fairness (ETS, 2002)[1]. Internal audits assess the compliance of each of its programs against these standards. Its staff has expertise and experience in a broad range of areas including: assessment design, measurement models and applications, data analysis and technology research, statistical theory and practice, college-readiness assessments, and international assessments.

ETS remains engaged with a broad range of research activities related to validity, reliability, score interpretation, assessment innovations, including new technologies used to develop and administer assessments, and accessibility and fairness of tests for all test takers. Population-centered validity research focuses on test validity for minority group members,

[1] *ETS standards for quality and fairness.* Princeton, NJ: Educational Testing Service.

women, and persons with disabilities. Documentation for criterion and construct validity of achievement tests is the subject of comprehensive review and evaluation.

Validity Research: Accessibility

The Foundational and Validity Research Center directed by Dr. Cara Laitusis conducts a wide range of validity research on existing assessment programs, including *GRE* and K–12 academic assessments used for accountability purposes. The center's work includes research on accessible assessments for students with disabilities, as well as universal design.

Dr. Laitusis is the principal investigator for the U.S. Department of Education-funded Designing Accessible Reading Assessments (DARA), which culminated with the development of a multistate reading assessment that provided direct measures of oral reading fluency and comprehension of text via audio. She also led a research team to investigate the accessibility issues of the new item types aligned with Common Core State Standards for students with disabilities and English learners. The PARCC consortium funded this research, and Dr. Laitusis's team provided consulting on the accessibility and APIP considerations for the consortium. ETS's research experience and expertise in this area will greatly benefit PARCC as ETS develops items and tasks that are accessible to all students. ETS proposes to continue its support of PARCC's accessibility features in both item development and delivery.

Program Management

ETS's project management capabilities, along with its quality assurance initiatives, are vital to the success of each task and activity in every program that the organization manages. The prime responsibility of the project management team is to make sure the program is delivered on time and on budget, and that all products and services delivered meet or exceed ETS's quality standards and PARCC's and Pearson's expectations.

The ETS project management group works with states, the Smarter Balanced and PARCC consortia, and other clients/programs ranging in size and complexity, from small item development programs to large, multifaceted programs involving numerous vendors and subcontractors. For example, in its work on the NAEP program, ETS serves as one of a number of contractors, each working independently yet coordinated as one seamless program. To accomplish this goal, the ETS program management staff works closely with NAEP staff at the U.S. Department of Education, as well as with the program management staff of the other NAEP vendors.

Within K-12, ETS serves in the role of subcontractor for programs such as the Smarter Balanced Assessment Consortium Item and Task Specifications contract, the Smarter Balanced Development of Accessibility and Accommodations Policies and Materials contract, the Tennessee Achievement Tests and End of Course Assessments, and the Mississippi Subject Area Testing Program; in the role of prime contractor for programs such as the

Smarter Balanced Psychometric Services contract; the Smarter Balanced Test Blueprint and Computer Adaptive Specifications contract, California Standardized Testing and Reporting Program and Maryland High School Assessments; and as one of several vendors for programs such as the Washington Comprehensive Assessment Program. For each of these programs, ETS has put into place a project management team and strategy specifically tailored for the requirements of the program.

The ETS project management team members assigned to PARCC have been specifically recruited for and assigned to PARCC based on ETS's understanding of your specific project management needs.

ETS's project management team knows that flexibility and creativity are very important to PARCC and its member states. ETS's project management team members have built excellent working relationships with PARCC, Pearson, and the many individuals involved in leading and supporting its work. Also, ETS has established a solid track record of collaborating with colleagues at Pearson as well as with representatives of the other contractors involved in the creation of the PARCC system. ETS is commited to continuing to listen and respond to your needs. All of ETS's program management activities are guided by its quality standards and by the industry's best practices as described in *Operational Best Practices for Statewide Large Scale Assessment Programs* (of Chief State School Officers and the Association of Test Publishers, 2010).

Experience Serving State Assessment Programs

ETS has developed large-scale criterion-referenced tests (CRTs) for statewide K–12 standards-based assessment programs since 2001. In addition, ETS has provided test development and psychometrics services for the NAEP since 1982.

ETS's experience encompasses the following:

- Assessment development for K–12 achievement (including alternate, modified standards, and simplified-English assessments for special populations), college and graduate school admissions, English language proficiency, and teacher licensure and certification
- Test design and measurement theory
- Psychometric analysis
- Educational and policy research
- Test development project management

ETS's current K–12 contracts include the following:

ETS Current Contracts		
California Standardized Testing and Reporting (STAR) California High School Exit Examination (CAHSEE) CAHSEE Alternate Assessment California Modified Assessment (CMA) California Standards Tests (CST) California Standards-based Tests in Spanish (STS) California Alternate Performance Assessment (CAPA)	Maryland High School Assessments (HSA) Maryland Modified High School Assessments (Mod-HSA) Mississippi Subject Area Testing Program, Second Edition (SATP2)	Partnership for Assessment of Readiness for College and Careers (PARCC) Item Development PARCC Assessment Administration
Smarter Balanced Assessment Consortium Item and Task Specifications, Style Guide, Bias and Sensitivity Guidelines, and Accessibility and Accommodations Guidelines Smarter Balanced Psychometric Services Smarter Balanced Development of Accessibility and Accommodations Policies and Materials Smarter Balanced Test Blueprint and Computer Adaptive Test Specifications, CAT Simulations and Interim System Specifications Smarter Test Administration for the Field Test	Texas Assessments of Academic Readiness (STAAR) Tennessee Achievement Tests Tennessee End of Course Assessments	Virginia Standards of Learning (VA SOL) Writing Assessments, including technology-enhanced items Virginia Standards of Learning (VA SOL) Grades 3–8 Assessments Virginia Standards of Learning (VA SOL) End-of-Course Assessments, including Technology-Enhanced items
Washington Measurements of Student Progress—Grade 3-8 Washington High School Proficiency Exam Washington End-of-Course Math & Science and Accommodated Forms	Wyoming Proficiency Assessment of Wyoming Students (PAWS) Wyoming PAWS Alternate Exam Wyoming Statewide Assessment of Student Writing	

ETS References

State	Maryland
Contract	**Maryland High School Assessment (HSA)** 2007–2012 (Five-Year Contract, 2012–2014 (Two-Year Extension), 2014–2016 (Two-Year Extension)
Overview	**Number of Students:** Currently 400,000 tested annually **Subject Areas:** Algebra/Data Analysis, Biology, English, Government **Grade Levels:** High school end-of-course **Test Type:** Paper and online
Products and Services Provided by ETS	**Overall assessment plan and tests:** The Maryland High School Assessments (HSA) are a series of end-of-course tests that cover core academic areas. The Algebra/Data Analysis, Biology, and English end-of-course tests fulfill the No Child Left Behind requirement for English, Mathematics, and Science at the high school level. The HSA series of tests included Government through the May 2011 administration before it was discontinued, however, government was reinstated with the January 2013 administration. ETS works collaboratively with Maryland educators in developing the overall high school assessment plan and the tests themselves. ETS also provides all these same services for a Modified High School Assessment. In May 2009, ETS and our subcontractor, Pearson Educational Measurement, became responsible for retrieving used and unused test materials, scanning student responses, and conducting secure check-in. **Responsibilities:** ETS is responsible for the following: (a) developing all test items and materials; (b) printing and delivering test books (including alternate formats in large print, braille, online audio, and Kurzweil), answer sheets, test administration manuals, examiner manuals, and ancillary materials for five test administrations per year (October, January, April, May, and Summer); (c) coordinating and facilitating summer committee meetings of Maryland educators; (d) designing and conducting psychometric analyses of test results, performing research studies, and producing an annual technical report; and (e) producing and delivering scores and score reports. Item types consist of selected-response (multiple-choice), student-response (grid-in) items, and constructed response in government beginning with the January 2014 administration. Online testing for all administrations of both the HSA and Modified exams was in place by May 2009. **Web-based ordering system:** We also offer schools and local education agencies (LEAs) a web-based ordering system. Schools and LEAs are able to go online to upload pretest files, order additional materials, order make-up materials, and check the status of orders.
Client Contact	Janet Bagsby Chief, Planning and Assessment Branches Maryland State Department of Education 200 West Baltimore Street Baltimore, Maryland 21201-2595 Phone (410) 767-0048 Fax: (410) 767-0100 E-mail: jbagsby@msde.state.md.us Website: www.marylandpublicschools.org/MSDE/

	National Assessment of Educational Progress (NAEP)
Contract	**National Assessment of Educational Progress (NAEP)** 1984–2001, 2002–2007, 2007–2018
Overview	NAEP, developed for the National Center for Educational Statistics of the U.S. Department of Education, is the only nationally representative and continuing assessment of the educational achievement of American fourth-, eighth-, and twelfth-graders. ETS coordinates an alliance of four other corporations that develop, administer, score, and report the results of the assessments. Since 1969, assessments have been conducted periodically in reading, writing, mathematics, science, U.S. history, geography, and other fields. Recent federal legislation has increased the visibility of NAEP's results, and all states now participate in its biennial reading and mathematics assessments. Nearly one million students were assessed by NAEP in 2013. ETS has developed solutions to NAEP's unique requirements as a nationally representative, sample-based assessment that covers a wide range of subject matter within short student assessment times while using both multiple-choice and constructed-response test questions. ETS develops the test question pools, and along with the client and the policy-setting agency (the National Assessment Governing Board (NAGB)), shepherds them through a series of reviews that include inspection by representatives of state education departments. NAEP provides results for the nation, for states, and for certain urban school districts for major demographic groups and, using student, teacher, and school background questionnaires, collects information about instructional programs and practices. NAEP has generated hundreds of reports across 12 subjects, providing information on students' academic performance, learning strategies, and classroom experiences.
Products and Services Provided by ETS	▪ Project coordination across a four corporation alliance ▪ Test development and scoring oversight ▪ Development of specialized psychometric analysis techniques ▪ Operational psychometric analysis, scaling, trend monitoring ▪ State-of-the-art reporting for multiple audiences in multiple formats ▪ Development of presentation materials for the client and conferences ▪ Development of web-based analysis tools open to the public ▪ Provision of content for the client's web site ▪ Consultation with national and state policy makers ▪ Support of the NAEP state services center– materials and instruction ▪ Client-requested psychometric research – both theoretical and project-related
Client Contact	Suzanne Triplett, Program Manager National Center for Education Statistics 1990 K Street NW Washington, DC 20001 Telephone: (202) 502-7465 E-mail: suzanne.triplett@ed.gov

	Smarter Balanced Assessment Consortium
Contract	**SBAC 04–Item and Task Specifications, Style Guide, Bias and Sensitivity Guidelines, and Accessibility and Accommodations Guidelines** December 1, 2001–April 16, 2012 **SBAC 05–Psychometric Services** February 2012–September 2014 **SBAC 06–Develop Accessibility and Accommodations Policies and Materials for the Smarter Balanced Assessment Consortium** January 2012–June 2012 **SBAC 09–Test Blueprint and Computer Adaptive Test Specifications, CAT Simulations and Interim System Specifications** February 8, 2012–May 31, 2012 **SBAC 15–Develop the Reporting System for the Summative and Interim Assessments** September 13, 2012–September 30, 2014 **SBAC 19b–Test Administration for the Field Test** June 13, 2013–September 30, 2014
Overview	**Subject Areas:** English Language Arts and Mathematics **Grade Levels:** Grades 3-8 and High School **Test Type:** Computer Adaptive Test
Products and Services Provided by ETS	**SBAC 04:** ETS provides a variety of psychometric services to the multistate consortium. Tasks include: ▪ Math General Specifications, including specifications for Selected Response, Constructed Response and Stimulus materials ▪ Math Item Specifications ▪ Math Sample Items (Technology Enhanced, Performance Task, Extended Response, Constructed Response and Selected Response) ▪ English Language Arts Item Specifications (Technology Enhanced, Performance Task, Constructed Response, Extended Response) ▪ English Language Arts Sample Items ▪ Bias and Sensitivity Guidelines **SBAC 05:** ETS (as a sub-contractor) provides the following services as part of the larger Item Specifications Project: ▪ developing a pilot test design ▪ developing technical manuals ▪ determining a standard setting design ▪ determining a vertical scale design ▪ designing a pilot test sampling plan ▪ developing pilot test item and task data review materials ▪ conducting pilot test analyses ▪ selecting anchor items and tasks for calibration and building the vertical scale ▪ verifying final field test forms ▪ conducting psychometric analyses to support field test data review ▪ conducting psychometric analyses to support item and task calibration ▪ presenting information and issues to the Technical Advisory Committee **SBAC 06:** ETS responsibilities as a subcontractor to Measured Progress included:

	Smarter Balanced Assessment Consortium
	▪ Component 1: ETS conducted a survey of consortium member states to determine current definitions, participation, and accessibility and accommodations guidelines and policies for English Language Learners (ELL), Students with Disabilities (SWD), Students with 504 Plans, and ELL SWDs. ETS also collected and reviewed literature and national data on accessibility and accommodations for the four subgroups. Tests accommodations and tools such as audio presentation of test content (read aloud), test translation (including ASL avatar and alternate language glossary), calculator use, and writing accommodations (the use of a scribe, spell check, and grammar check) were researched. A final report on the methodology and findings was prepared. ▪ Component 2: A case study on tactile representation in assessment was conducted and reported. Visually impaired students in Oregon were observed and post-test interviews were conducted with the students and their parents. ▪ Component 3: Senior ETS researchers will review an accessibility and accommodations manual prepared by Measured Progress and participate in dissemination events with Smarter Balanced stakeholders. **SBAC 09:** Overall Responsibilities for ETS as the Prime Contractor for this four month contract: ETS staff will lead and collaborate with the Consortium, Consortium selected Expert Advisors, and consultants to attain these objectives. ▪ Component 1: Develop test specifications and blueprints. ○ Develop statement of test purpose and develop specific design for representing Smarter Balanced's content specifications for summative test. ○ Develop overall test specifications including test length, psychometric criteria and the number of times students are tested from the same pool ▪ Component 2: Develop pilot and field test specifications ○ Develop pilot test form specifications with intended difficulty distribution. ○ Develop field test specifications including anchor test design ▪ Component 3: Develop CAT specifications and conduct initial CAT simulation studies ○ Develop algorithm for computer adaptive tests. ○ Provide a simulation engine that executes algorithm and carry out simulations for each test, varying pool and population characteristics as well as CAT constraints. ○ Conduct simulation studies. Based on simulation results, recommend optimal design, pool structure, and pool distribution. ○ Provide test level simulation output and reports of student parameter recovery. ○ Develop progress reports on CAT simulation studies and identify optimal algorithm design. ○ Provide documentation and consultation regarding recommendations to address federal RTTA monitoring and peer review standards ▪ Component 4: Develop specifications for an interim testing system to include, at minimum, a collection of items and tasks and an interface for allowing states and districts to craft assessments to inform instruction ○ Create specifications for the purposes and functions of the interim system by facilitating discussions among SBAC stakeholders. ○ Specify a framework for the bank of items and tasks and the interface

	Smarter Balanced Assessment Consortium
	for state and district use. **SBAC 15:** ETS (as a sub-contractor to Wireless Generation) provides the following services as part of the larger Reporting System Project: ▪ Reporting Systems Requirements where ETS to advise and propose an approach to Wireless Generation (WGen) for best practices to gather requirements for reporting. ○ Suggest participant make-up for the requirements gathering sessions ○ Review and feedback on the WGen final design ○ Share sample reports from other ETS sources ○ Review and provide feedback on the WGen analysis of the data from requirements gathering sessions ○ Provide advice and collaborate with WGen to assure requirements are tied to the Smarter Balanced Theory of Action. ▪ Report Design where ETS will provide report design consultation and review/feedback to WGen throughout the report development process with respect to: ○ Psychometric validity ○ ELL populations ○ Accessibility and Accommodations ○ Common Core Standards – traits, claims, and blueprints ○ Meeting minimum standards to technical quality and adherence to national assessment standards ▪ Report Deployment and User Acceptance Testing (UAT) ○ Create Sampling Plan designs for the Beta UAT ○ Create Sampling Plan designs for the Large Scale UAT **SBAC 19b**: This contract provides support for the Smarter Balanced Field Test Administration and supporting documentation for the Operational Assessment. ETS is required to satisfactory execute on 25 deliverables, related to administration of the field test and/or operational test. ▪ Administration Manuals (FT and Operational) ▪ Training Modules (FT and Operational) ▪ Help-Desk Support (FT only) ▪ Recruit/Register FT participants (FT only) ▪ Communication Strategies (FT only)
Client Contact	Michael Middleton Director, Office of Superintendent of Public Instruction Old Capitol Building P.O. Box 47200 Olympia, WA 98504-7200 Phone: (360) 725-6434 Fax: (360) 725-0424 Email: Michael.Middleton@k12.wa.us Website: http://www.k12.wa.us

About Measured Progress

Corporate Overview and Capabilities

Measured Progress is a full-service, customized, standards-based general and alternate assessment contractor providing consortium-, state- and district-level a complete spectrum of assessment services. The company was originally incorporated 30 years ago as Advanced Systems in Measurement and Evaluation, Inc. Measured Progress began with a staff of four who worked in a small suite in historic Portsmouth, New Hampshire. The company assumed its current name and not-for-profit 501(c)3 status in 2000, and it has grown to nearly 500 full-time staff. Measured Progress supplements this number with seasonal temporary staff to meet contractual obligations as required.

Their corporate headquarters are located in Dover, New Hampshire. The main campus includes a 100,000 square-foot corporate office building, a 74,000 square-foot processing center, and a 36,000 square-foot office of information technology. In addition to their corporate campus, Measured Progress also operates facilities in Lee, New Hampshire; Newton, Massachusetts; Redding, California; Longmont, Colorado; and Menands, New York. These locations provide additional capacity to Measured Progress operations.

Measured Progress Philosophy

As an organization committed to educational assessment, Measured Progress believes its corporate and educational values should be consistent. Therefore, underlying their philosophy is the fundamental assumption that all individuals can achieve at high levels, especially when they understand what is expected of them and they operate in a supportive environment. In education, the goal is high academic performance; at Measured Progress, the goal is to provide high-quality products and services that make an impact. To achieve these goals, either in the classroom or the workplace, a collegial, collaborative, productive environment is a necessity. They strive to create assessment programs that meaningfully connect their clients' assessments with what is happening at the classroom level.

Central to their philosophy is the principle that accountability for quality is shared. In our schools, administrators, teachers, parents, and students all have responsibility for the quality of teaching and learning. At Measured Progress, all employees have a sense of mutual ownership and pride in our products and services, and share the responsibility for the quality of our work.

Also critical to the quality of their work are the strategic hallmarks of client service for which we are known—responsiveness and flexibility. Just as educational programs should be responsive to the individual needs and learning styles of students, Measured Progress is responsive to the unique needs of each client. Similarly, employees share an appreciation of divergent needs and work styles.

This understanding helps Measured Progress to best accomplish its mission and vision, and allows them to continue to serve education and, ultimately, the students.

Staff Resources

A significant percentage of Measured Progress' staff members are former educators, including many of our senior managers, and many have advanced degrees. Employees play an active role in scholarly and collegial activities, frequently presenting papers and presentations to trade and educator groups, and are veteran researchers and presenters considered to be at the forefront of their respective fields. They also serve on association and advisory panels, advise legislators and other education policy makers, and serve as a resource to national media.

Overviews of Measured Progress' key areas of functional expertise and experience related to the scope of this proposal are provided below.

Accessibility Services

Measured Progress is a leader in the development of the Accessible Portable Item Protocol (APIP) Standard—an interoperability standard for test item content and for storing student access needs.

The Measured Progress Innovation Lab specializes in innovative approaches to accessibility and accommodations for students with disabilities and special needs. The Innovation Lab worked with the IMS Global Learning Consortium and eight states, spearheaded by the Minnesota Department of Education, to develop APIP – allowing for test items and associated accessibility information to be ported between systems. APIP is currently maintained by the IMS Global Learning Consortium as an open-source, open-license standard.

APIP enables test delivery engines to tailor the presentation of items to meet individual examinees' access needs. The APIP development team sought to provide assessment programs and item developers with a structure for standardizing the file format of digital test items that could also be used to specify all the information and resources required to make a test item accessible for students with a variety of disabilities and special needs, as well as English language learners.
The specifications within APIP cover three critical areas:

- Content: The accessible content, which provides a wide range of accessibility supports for digital test content.
- PNP: The user Personal Needs and Preferences profile, which documents the specific accessibility needs of each student.
- Delivery: The delivery system, which combines the user needs with the content so that the item is accessible for the student.

APIP provides the structure, format, and language to specify the exact manner in which tailored representations are to be provided by a test delivery system for a student's specific needs. Now that this technical solution exists, in order for computer-based testing platforms to deliver high-quality alternate representations of test content in a standardized and equitable manner, a set of guidelines for how to appropriately represent content in different forms is essential. The Innovation Lab has conducted a series of projects to meet this need.

The Measured Progress staff continues their research and development, including conducting formal efficacy studies, related to computer-based/online accommodations, and explores new assessment approaches, including

- development of computer-adaptive testing within the accommodated delivery system;
- development of technology-enhanced, innovative items;
- development of diagnostic assessments that are driven by learning progressions and designed to identify common misconceptions; and
- the potential use of a variety of devices (e.g., mobile devices) to deliver curriculum and assessment content

Innovation Lab staff members are actively engaged in two work groups designed to address APIP implementation issues. The APIP End User Group, facilitated by IMS Global, consists of vendors and state department of education representatives. The APIP Work Group, also facilitated by IMS Global, includes several state and testing industry members, including Measured Progress, Pearson, ETS, CTB/McGraw-Hill, ACT, and CAL.

Research and Analysis

The Measured Progress Research and Analysis (R&A) department, made up of data analysis, psychometric, and quality control staff, uses a systems approach to compute, analyze, and program both score and demographic data for standard setting, test item analysis, and special studies. R&A staff also perform scoring, scaling, and equating analysis to produce student-level and aggregated reports. Their procedures promote efficiency, accuracy, high standards, and security in every phase of our work. Statistical analyses include built-in quality control checks and redundant cycles to verify data accuracy.

During all phases of a project, program management and technical staff compile documentation that supports the quality of our services and products. An important purpose of documentation is the replication of the analyses in subsequent years. In support of their focus on quality, the Measured Progress Technical Advisory Committee (TAC) reviews and advises on all contract designs.

Measured Progress psychometricians possess experience and expertise in a broad array of psychometric models addressing a host of issues from standard setting and equating to conducting comparability and validity studies. The outcome is an ability to apply the best solutions to meet each client’s unique needs. Large-scale assessment programs that report the results of their assessments by proficiency levels often require standard-setting studies to determine the threshold of total test scores separating the proficiency levels. They are experienced in all models of standard setting and have conducted standard-setting studies for large-scale assessment programs in many states including Colorado, Florida, Maine Massachusetts, Missouri, Montana, Nevada, New Hampshire, New Jersey, New Mexico, New York, Rhode Island, and South Carolina.

Program Management

Keeping projects successfully managed and on time is critical to the success of any assessment program. Measured Progress program management staff members act as primary liaisons between the Measured Progress organization and our clients. Staff members schedule, oversee, and facilitate each phase of a contract, verifying that all test development, production, and distribution work is completed on time and in accordance with contract terms and conditions. Program managers and program assistants work very closely and have frequent communication with clients to keep them apprised of the status of all program activities. They maintain consistent contact with our clients, and develop customized communication plans so that our staff is always accessible. Measured Progress program managers strive to maintain transparency with our clients, and if there is an issue that needs to be escalated, senior management is accessible and available to assist in whatever capacity necessary.

References for Measured Progress

State	Rhode Island
Contract	**Rhode Island Interim Assessment** 2011–2014
Overview	**Assessments Delivered:** Interim assessments in fall, winter, spring of each contract year in ELA and math, grades 3-11. Test building engine for users to create their own tests **Students Tested**: Up to 10,000 per grade, depending on the number of districts choosing to use the system. The Rhode Island Interim Assessment system is offered to interested districts in Rhode Island. The online system will allow users to create assessments; schedule tests; administer tests to students online or via paper and pencil; scan paper responses back into the online system; and score open-response items. The Rhode Island Department of Education and Measured Progress worked together to develop a set of fixed-form interim assessments designed to measure students' progress throughout the school year. Additionally, a test building engine will allow teachers to create their own tests using items aligned to the Common Core State Standards.
Products and Services Provided by Pearson	▪ Program management ▪ Item development aligned to the common core ▪ Development of a technology platform for creating, delivering, and scoring assessments

State	Rhode Island
	▪ Training ▪ Technical support
Client Contact	Jessica Bailey, Assessment Specialist Rhode Island Department of Education Office of Instruction, Assessment, & Curriculum 255 Westminster Street, 4th Floor Providence, RI 02909 Phone: 401-222-8253 Email: jessica.bailey@ride.ri.gov

	National Center and State Collaborative – General Supervision Enhancement Grant Project
Contract	**Item Writing Project** 2012–2014
Overview	**Subcontractor:** Questar Assessment, Inc. The NCSC Item Writing Project is an item development contract. Measured Progress, along with subcontractor Questar, has developed mathematics and English Language Arts items for the NCSC summative assessment. This will be a summative alternate assessment for students with significant cognitive disabilities. NCSC currently has 15 partner states and 11 Tier II states. The items have been developed in item family sets that are made up of 4 items each, where the Tier 4 item is the most cognitively complex and the Tier 1 item the least. The items have been written so that they may be used for both paper and computer based assessments. The items are graphics and teacher directive intensive.
Products and Services Provided by Pearson	▪ Program management ▪ Item development aligned to the common core through the Common Core Connectors ▪ In-person training, and facilitation of item bias and review committee meetings ▪ Development of alternative text tags for item graphics as appropriate ▪ Delivery of APIP compliant QTI
Client Contact	Sharon E. Hall, Ed.D. Director, NCSC Assessment Systems Principal Associate edCount, LLC 5335 Wisconsin Avenue Suite 440 Washington, DC 20015 Phone: 202-400-0909 Email: shall@edcount.com

	Partnership for Assessment of Readiness for College and Careers (PARCC)
Contract	**PARCC Item Development** 2012–2015
Overview	**Measured Progress Program Director and Manager:** Cathy Schirmer The PARCC Development Team — comprising ETS, Measured Progress, and CTB/McGraw-Hill — is responsible for the design and development of the items and tasks for the Mid-Year, Performance-Based, and End-of-Year/End-of-Course assessments across grade levels and components (English Language Arts/Literacy and Mathematics). Using Measured Progress' APIP Services Tool, Measured Progress processed over 7500 passages and items for PARCC.
Products and Services Provided by Pearson	▪ APIP Services ▪ Item Development ▪ Program Management
Client Contact	Kit Viator Executive Director, K-12 Multi-State Assessment Solutions Educational Testing Services 660 Rosedale Road Princeton, NJ 08541 Phone: 508-479-6840 Email: kviator@ets.org

About WestEd

Corporate Qualifications and Management Support

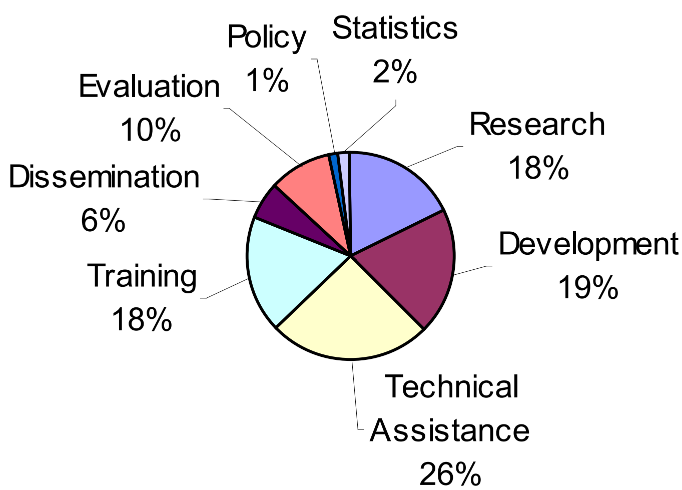
History and Governance Structure of WestEd

WestEd is a Joint Powers Agency (JPA), authorized in 1995 by a California Joint Powers Agreement and governed by public entities in Arizona, California, Nevada, and Utah, with Board members representing agencies from these states and nationally. WestEd's two predecessors, Far West Laboratory for Educational Research and Development (FWL) and Southwest Regional Laboratory (SWRL), were JPAs created in 1966. Since 2000, WestEd has carried out more than 4,000 successful projects representing major contributions to the nation's research and development (R&D) resources, and has more than 400 active contracts at any given time.

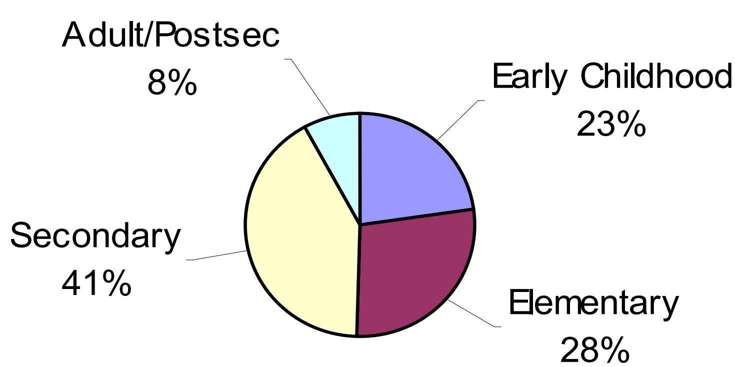
Current work extends beyond the western region to include most states in the nation and an increasing number of other countries. In Fiscal Year (FY) 2013, the agency is expecting to operate on program funding of approximately $125 million. Funding for specific projects comes from sources including the U.S. Department of Education (ED), the National Science Foundation, and the U.S. Department of Justice; state departments of education; and universities, school districts, foundations, and state and local agencies across the country.

WestEd has been vetted and approved as a qualified service provider in the U.S. Department of Health and Human Services Program Support Center (PSC) Task Order Contracts and the General Service Administration's Mission Oriented Business Integrated Services (MOBIS) Schedule federal contracting programs. This large variety of funding sources provides WestEd with a stable funding base and thus a stable organizational structure for achieving the work of this proposal.

WestEd's mission—to work with education and other communities to promote excellence, achieve equity, and improve learning for children, youth, and adults—is addressed through a comprehensive range of projects. The first figure, Work Type of R&D Activity, indicates how WestEd's core work is distributed by type of R&D activity (based on the FY 2012 project characteristics survey). The second figure, Work Target Level of Education, displays the distribution by target level of education served.



WestEd Work Type of R&D Activity (2012)



WestEd Work Target Level of Education

To carry out this mission, WestEd project staff are organized into a dozen formal program areas—some addressing educational content or level and some in areas of high risk and high need that cut across content. Areas of work include school and district improvement; early childhood; mathematics and science; English learners; assessment and accountability; and evaluation.

Across programs, WestEd boasts expertise in student assessment, data-driven planning, curriculum development, training, school coaching, community-partnership building, research and evaluation methods, and policy analysis. Collaboration among staff is institutionally promoted through regular meetings of the management, program, and administrative councils.

Corporate Organization and Resources

WestEd is governed by a Board of Directors representing the four western region states of Arizona, California, Nevada, and Utah, and is directed by the agency's Chief Executive Officer, Dr. Glen Harvey. The agency currently employs 605 regular professional, support, and administrative staff. WestEd staff hold at least 379 advanced degrees, including 123 doctorates in education or fields such as psychology, sociology, public policy, statistics, and law. Most staff have years of experience in research, development, staff training, technical assistance, evaluation, and/or policy activities. Many members of the senior staff are known nationally for their work in their fields. Their reputation and achievements are recognized by awards from professional organizations, placement on boards, and selection for high-profile advisory committees.

Daily business operations—including contract administration, contract compliance, data processing, and accounting functions—are handled through WestEd's Contracts, Accounting, and Finance departments, under the direction of Nancy Riddle, Chief Financial Officer (CFO). Key positions include the Director of Finance, Director of Legal and Compliance, Director of Contracts Management, Contracts Administrator, Controller, Accounting System Coordinator, Accounts Receivable Manager, Accounting Operations Supervisor, and Compliance Officer. Practices are governed by standard accounting principles, the agency's Rules for the Conduct of Business, the rules governing government contracts, and specific contractual agreements. The majority of the agency's contracts are billed on a cost-reimbursement basis. Accounting, billing, and reporting procedures are designed specifically to meet a variety of government reporting requirements, such as FAR, EDAR, and EDGAR.

Assessment Development

WestEd has worked with states, districts, and local schools to gauge student performance and make the best use of achievement results. Its breadth of knowledge in the area of assessment and test development is extensive. WestEd has developed large-scale assessments for numerous states across the country, including Arizona, California, Kansas, Kentucky, Louisiana, Massachusetts, Nevada, Pennsylvania, and West Virginia.

WestEd's Assessment Experience. In the following table we provide an overview of the variety of services that WestEd's Assessment and Standards Development Services (ASDS) program has offered since 2002. In addition to item and test development, ASDS has experience in standards review and development, conducting alignment studies, consultation on special populations, facilitation and/or participation in Technical Advisory Committees, and technical and/or policy support and consultation. As indicated by the diversity of work listed in the following figure, the WestEd team assigned and bid for this project has diverse experience in supporting statewide large-scale assessment programs across the nation.

Services that WestEd's Assessment and Standards Development Services						
State Educational Agency	**Standards Review and Development**	**Item and Test Development**	**Alignment Studies**	**Special Populations**	**Technical Advisory Committee**	**Technical/ Policy Consultation**
Alaska					X	X
Arizona	X	X	X	X		X
Arkansas			X			X
California	X	X	X	X	X	X
Colorado	X		X		X	X
Georgia		X	X			X
Idaho	X			X	X	
Iowa			X	X		
Kansas		X		X	X	X
Kentucky		X	X	X		X
Louisiana	X	X	X	X	X	X
Maine			X	X		
Maryland		X				X
Massachusetts		X		X		
Montana				X	X	X
Nevada	X	X	X	X	X	X
New Hampshire	X			X	X	
New Jersey			X			X
New Mexico	X	X	X	X	X	X
New York					X	X
Ohio		X			X	
Oklahoma				X		
Oregon	X		X	X	X	X
Pennsylvania		X			X	X
Rhode Island			X		X	
South Carolina						X
Utah	X	X	X		X	X
Vermont	X		X		X	
West Virginia		X				

Services that WestEd's Assessment and Standards Development Services						
State Educational Agency	**Standards Review and Development**	**Item and Test Development**	**Alignment Studies**	**Special Populations**	**Technical Advisory Committee**	**Technical/ Policy Consultation**
Contractor for PARCC States	X	X	X	X		
PARCC		X				
Project Management Partner (PMP) Smarter Balanced	X	X	X	X	X	X

WestEd Assessment and Standards Development Services Experience from 2002–2013

WestEd References

Contract	**PARCC Item Development–Measured Progress**
Period	September 12, 2012-Current
Overview	As subcontractor to Measured Progress, WestEd provided item development services to support the development of the PARCC assessments. WestEd wrote technology-enhanced items for grades 7, 8, and 9 in ELA and grades 6, 7, and 8 in mathematics
Client Contact	Tim Crockett, Senior Vice President, Measured Progress 100 Education Way, Dover, NH 03821 (603) 749-9102 Crockett.Tim@measuredprogress.org

Contract	**PARCC Item Development–Pearson**
Period	June 1, 2013–March 31, 2015
Overview	As subcontractor to Pearson, WestEd is providing item development services to support the development of the Partnership for Assessment of Readiness for College and Careers (PARCC) assessments. WestEd is developing items for grades 3–8 in both ELA and mathematics.
Client Contact	Margaret Kramer, Director, Measurement Development Services, Pearson Inc. 2510 N. Dodge St., Iowa City, IA 52245 (319)339-6736 Margaret.kramer@pearson.com

Contract	Support Services for Nevada Student Assessments
Period	July 1, 2010–June 30, 2015
Overview	WestEd serves as the test development contractor for Nevada's Proficiency Examination Program, which includes criterion-referenced assessments of reading and mathematics for grades 3–8 and science for grades 5 and 8, and high school proficiency examinations for reading, mathematics, and science. WestEd is responsible for item development, content and bias review meetings, test form construction, development of camera-ready test forms, scoring guides, and identification of anchor and training papers to support the scoring of student work.
Client Contact	Cindy Sharp, Director for the Office of Assessment, Program Accountability and Curriculum (APAC) 700 E. Fifth Street, Carson City, NV 89701 (775) 687-9166 csharp@doe.nv.gove

Component 1: Test Development

V.A.1. Item Development

Requirement

V.A.1.A. PARCC Test Development Documentation

1. PARCC Blueprints
2. Supporting Documentation
3. Performance Level Descriptors

Deliverables for Section V.A.1.A.

a) Contractor will maintain PARCC assessment specifications documents as needed or required, and deliver updated documents each year. Documents to be maintained include:

i. ELA/Literacy and Mathematics Item Specifications (which include Evidence Statement Tables)
ii. ELA/Literacy and Mathematics Form Specification Tables
iii. ELA/Literacy Task Generation Models
iv. Mathematics Design Patterns
v. High School Course Level Specification Documents for Mathematics
vi. PARCC Model Content Frameworks for ELA/Literacy
vii. PARCC Model Content Frameworks for Mathematics
viii. PARCC Style Guide
ix. Accessibility Guidelines
x. Cognitive Complexity Measures
xi. Text Complexity Measures
xii. Linguistic Complexity Measures
xiii. PARCC Accessibility Features and Accommodations Manual
xiv. PARCC Performance Level Descriptors

Response

At the beginning of the contract, we will work with PARCC to verify that we have a mutual understanding of test development requirements, including those described in the PARCC Test Development Documents.

As an initial starting point, PARCC will deliver to Pearson complete and up-to-date PARCC blueprints, supporting documentation, and performance level descriptors as specified in the

RFP within five days of contract execution. After these documents are reviewed, Pearson and its subcontractors will determine if any additional questions or clarifications are needed. Otherwise, they along with the solutions presented in this proposal, will be the basis for the development of all new content under this contract.

Each year, Pearson and its subcontractors will update the PARCC blueprints, supporting documentation, and performance level descriptors are specified in the RFP and hand off to PARCC on an agreed upon schedule.

Requirement

V.A.1.B. Technology Requirements for Test Development

1. The PARCC Item Bank
2. Metadata
3. Item Encoding
4. APIP Metatagging
5. Interoperability Conformance and Validation
6. Technical Integration Requirements

Response Requirements for Section V.A.1.B.

a) Offeror's proposal shall include a response to the requirements specified in Section V.A.1.B.

Deliverables for Section V.A.1.B.

a) Contractor will maintain PARCC technical assessment specifications documents as needed or required, and deliver updated documents each year. Documents to be maintained include:
   - i. Metadata Schema
   - ii. PARCC Item Development Technical Guide
   - iii. PARCC interoperability and audio-visual guidelines
   - iv. Item Import/Export Quality Plan

Response

PARCC Item Bank

To support the technology requirements for the PARCC item bank during the operational test development activities, Pearson proposes using our next generation assessment banking and test building tools specifically designed to support the most modern standards-based interoperable assessment delivery platforms.

Our Assessment Banking and Building tools for Interoperability (known as ABBI) is a full suite of end-to-end assessment creation and management tools. For the PARCC operational test development activities, we will bring to bear the specific modules within ABBI that support the authoring and content export capabilities described by the PARCC test development technology requirements.

Usable on a Variety of Test Delivery Platforms

ABBI is forged for interoperability, which enables item content and metadata to be used across a variety of test delivery platforms that adhere to interoperability standards. This section will outline the ABBI solution and how it will meet the PARCC item bank technical requirements in support of the operational test development process.

ABBI is our latest addition to the next generation suite of systems that includes but is not limited to TestNav, ePEN, and PearsonAccess. ABBI development began in earnest in August 2013. While Pearson believes that ABBI is the right solution for the PARCC program and will satisfy all PARCC requirements over time. Pearson will begin transitioning PARCC content development activities to ABBI in 2014 starting with all ELA content being authoring directly in ABBI in September. This will include TestNav previewing capability built in from the start. Math will remain on Pearson's legacy systems as content development starts in 2014. Later in 2014, we will begin loading the new Math content into ABBI to support the content review cycles occurring early in 2015. In 2015 we will continue to build out the authoring tools to support the complex Math item types as well as enhancing the review, security, and workflow systems..

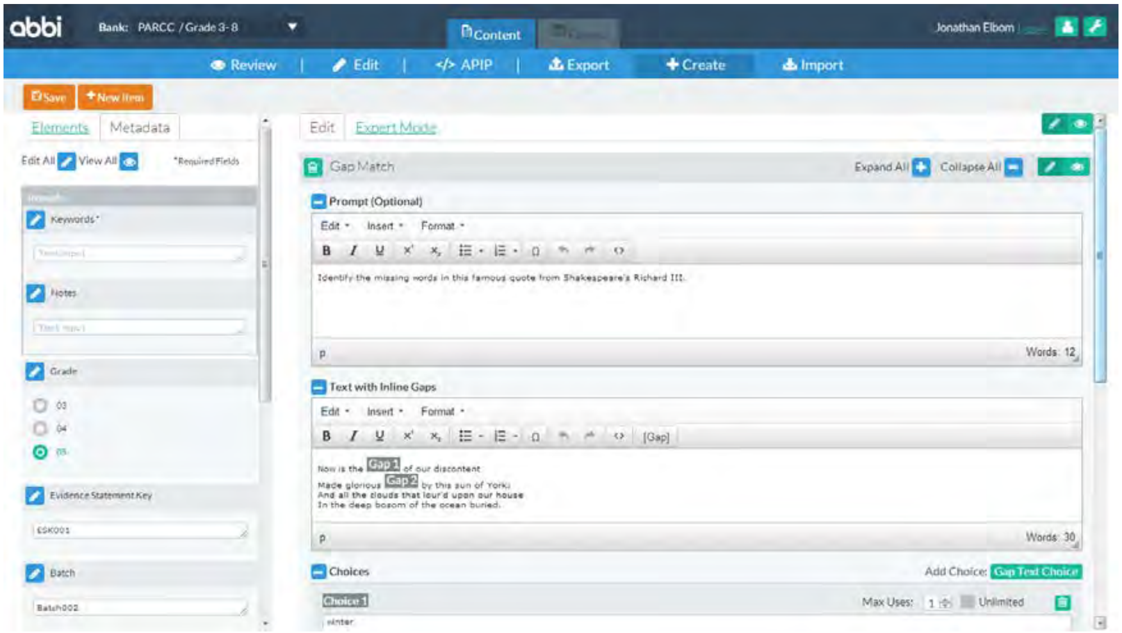
As new capabilities are added to ABBI, we plan to expose those capabilities and transition development activities to ABBI as early in the development cycle as possible and practical. While the transition to a full ABBI solution occurs, we will continue to use the existing capabilities in use for all prior phases of PARCC content development. Both capabilities can live in parallel and are fully supported. As any transitions occur, they will be thoughtfully planned and coordinated with all stakeholders.

A Secure Web-Based Application

The ABBI authoring system is a secure web-based application that authorized users can access using a standard web-browser that is HTML5 capable. All Internet communications use a secure encrypted transmission with Secure Sockets Layer (SSL) protocol. ABBI provides a standard role-based access control mechanism that will limit users to specific content, data, and functionality based on the role assigned.

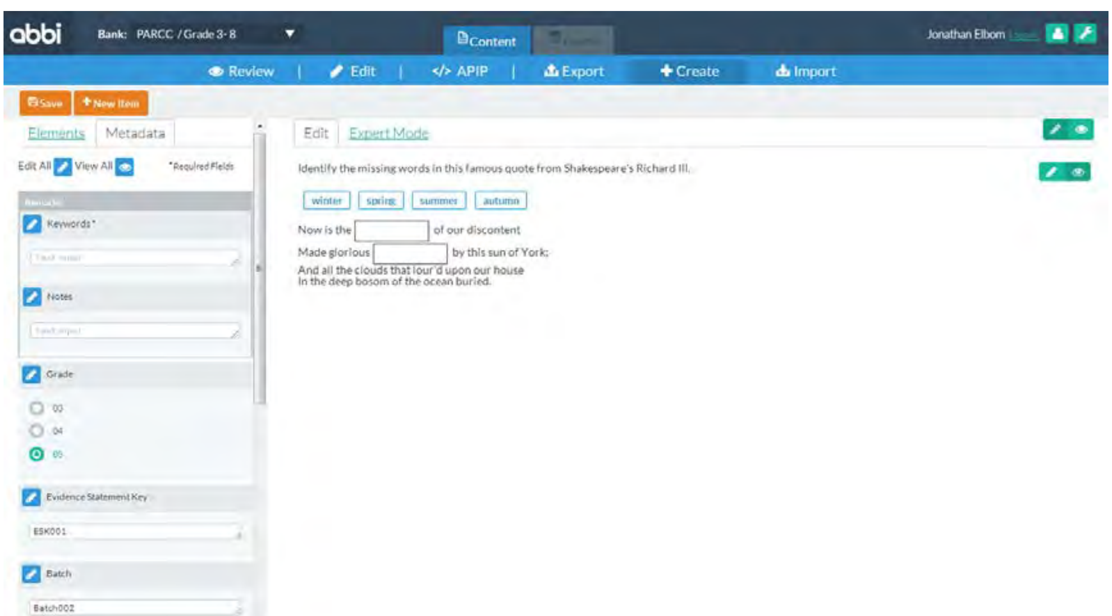
After a successful login, users will be presented with an intuitive and easy-to-navigate interface for viewing and managing the items and tests within the application based on their role. The ABBI banking capabilities will track all revisions to an item as it is edited.

The following figure shows how the authoring interface would appear to a user creating a standard gap match (i.e. drag-and-drop) item using the ABBI interface. A traditional text editing interface will allow the user to create the question and place gaps within inline text where the gap choices can be placed. Intuitive interfaces for creating choices and specifying correct responses are also provided to verify that the item is fully encoded.



ABBI Authoring Interface. The authoring interface as it would appear to a user creating a standard gap match (i.e. drag-and-drop) item using the ABBI interface.

The ABBI authoring tool will provide a quick view capability by clicking the green "eyeball" button in the upper right portion of the screen. The screen below illustrates this view and, as you can see, it removes all of the editing tools and displays just the content.



Viewing Tools and Content. The ABBI authoring tool will provide a quick view capability by clicking the green "eyeball" button in the upper right portion of the screen.

The underlying content model for the ABBI bank and the ABBI authoring interfaces are completely compatible with QTI and APIP specifications that are being implemented for the PARCC assessment programs. In fact, we are enabling all of the specific PARCC style attributes and interactions as part of the ABBI authoring design interfaces as those specifications are defined.

Items will be encoded using QTI 2.1 and APIP 1.0 specifications, with the assumption that accessibility features and student tools unrelated to response capture will be built into the test delivery platform. The authoring tools will also allow users to view and update PARCC item metadata at the same time they are viewing and updating the item. The metadata will be configured to PARCC specifications. ABBI supports the ability to secure specific metadata attributes to specific user roles if that level of control is desired.

The IMS QTI/APIP custom interaction and portable custom interaction specifications provide multiple methods for expanding the range of item interactions. Pearson will continue to help pioneer new QTI item models that effectively measure the complexity and rigor of the Common Core State Standards (CCSS) and verify that industry specifications are adapted appropriately to accommodate those interactions. This process will incorporate Partnership approval for any extensions to the PARCC data model and thorough documentation of those extensions.

Standardization and Adherence to the PARCC Data Model

By using ABBI authoring for item development we will achieve a very high degree of standardization and adherence to the PARCC data model. In addition to the standardization achieved through use of the ABBI authoring tool, ABBI's export routines will include schema validation against QTI/APIP as well as a stricter validation against a schema-based representation of the PARCC item content data model.

The Pearson content development team will obtain and identify all required permissions in the PARCC metadata fields for copyright and licensing. All items will be appropriately permissioned for use on PARCC assessments or released publicly through the Partnership Resource Center. In 2015, all PARCC items will be migrated from the existing item bank into Pearson’s proprietary system, ABBI, which will be the PARCC item bank during the term of the contract,for storage of all items, metadata and other content assets.

Pearson will provide PARCC with released items for states to access from the Partnership Resource Center (PRC). It is important to recognize that any test delivery system that adheres to interoperability standards should be able to consume PARCC content very effectively. However each delivery platform may have varying conventions for styling and formatting. These variations may require the receiving system to perform additional validations and transformation that are beyond the scope of this proposal. Pearson will work with PARCC and the PRC provider to verify that all content specifications are clearly documented, and Pearson will be available to answer questions about those specifications as needed. Pearson can make sample content packages of PARCC practice tests available to PARCC and the PRC provider to allow PARCC to verify the PRC is setup properly.

Metadata

PARCC's current schema for metadata has been carefully considered and synchronized with several existing standards. Pearson has a thorough understanding of this schema and experience working with it. ABBI's authoring capabilities allow for metadata to be configurable to the PARCC specification and for the metadata entry interfaces to be restricted to valid data. As the PARCC metadata schema or the supporting industry standards evolve over time, Pearson will work with the Partnership to incorporate the necessary changes as they are required.

In addition, opportunities may arise for improved alignment between metadata and APIP functionality (e.g., whether the item is appropriate for use with certain populations). Pearson is eager to work with the Partnership and IMS Global on further refinement of these data models to best deliver on the APIP model.

Item Encoding

ABBI will support all current and known PARCC item types as they are encoded in QTI and APIP. Items will be encoded using PARCC implementation of the QTI 2.1 and APIP 1.0 specifications. As the PARCC content model or the supporting industry standards evolve over time, Pearson will work with the Partnership to incorporate the necessary changes as they are required.

The ABBI content model and authoring tools are easily extensible to support new item types as the PARCC content requirements change. The ABBI authoring system will constrain content developers to the specific specification and vocabularies unique to the PARCC content model. This will result in the highest quality and most consistent and portable content possible.

The ABBI authoring interface will provide a method to preview the content in TestNav as the content is being authored. This will allow the author to verify that the content is presented properly in the TestNav platform. There will be some limitations of this previewing capability, such as accessing tools (rulers, calculators, etc.), full accessibility features, and shared content assets (such as passages) that may not be present when the preview request is made by the author. The TestNav previewer will be updated in future releases and some limitations may be overcome.

APIP Metatagging

ABBI includes two main components that support APIP tagging:

- The first component is a set of automated routines that can take a standard QTI item (without APIP tags) as input and add the APIP tags using program-level preferences for inserting identifiers and segmenting content into access elements as required.
- The second component is an intuitive user interface that allows the user to review the APIP extensions by inclusion order and access element known as the ABBI-APIP Editor. The user can manage the access elements to change text in support of text-to-speech pronunciations, spoken text, and braille. The user can preview the text-to-speech using an on-board engine that is integrated with ABBI.

If the base QTI contained alt tags for graphical content (a standard HTML structure), those tags can be carried forward to the APIP extensions and are available through the APIP editor user interfaces. Convenient tools for propagating changes to the alt tags to all associated access elements are available.

While it will be our recommended practice that APIP tags are added towards the end of the content development cycle, we will always emphasize accessibility first through our Universal Design and accessibility review processes. APIP tagging introduces an increased level of content maintenance that will largely be manually controlled after the APIP tags are added. Introducing APIP tags early in the life of an item while edits are frequent will increase the costs of development if APIP extensions have to be managed in parallel.

Pearson is committed to using the IMS Global APIP open interoperability standard for all assessment content and meta-data encoding. APIP provides a robust content tagging and meta-data vocabulary that is specifically designed for accessible assessment content encoding. APIP content tagging and extensions are designed to support a wide range of disabilities and capabilities within the delivery platform.

Pearson is currently building the authoring and banking tools required to manage APIP content as part of our ABBI development effort. The ABBI tools will consist of automated scripts for generating APIP tagging extensions to the base item content as well as providing an intuitive user interface to manage and maintain the accessible content.

Interoperability Conformance and Validation

Pearson will use appropriate data format validators to confirm that data items being transferred use appropriate open standard interoperability formats according to specifications. We anticipate the system components will present web services for transactions that employ industry standard data formats and that the PARCC test harness will provide a way to integrate its functionality into the data transfer process. These data formats will be based on SIF, QTI, and APIP with the goal of allowing external systems to interoperate based on common data representations.

Technical Integration Requirements

Pearson will be responsible for maintaining all system components and existing content that relies upon the underlying content encoding model. If the content model changes as a result of new features within Pearson's suite of next generation systems, those changes will be provided by Pearson. Any changes that are introduced by other parties for new capabilities or content requirements will be made after specifications have been reviewed and impacts assessed.

When items have been identified for release, Pearson will export standard QTI content packages from the PARCC item bank to PARCC or its designated recipient such as the Partnership Resource Center

Requirement

V.A.1.C. Item Development Planning and Targets

1. Item Development Strategy Meetings
2. Item Development Targets
3. Item Types
4. Technology-Enhanced Items
5. Paper-Pencil Items

Response Requirements for Section V.A.1.C.

a) The Partnership has experienced that the costs for permissions can range up to $10,000 as well as be difficult to obtain. The Partnership is seeking recommendations on how to contain costs as well as obtain all needed permissions.

b) Proposed distribution of TE's across item types for each content area and grade level

Deliverables for Section V.A.1.C.

a) Items, tasks, texts, stimulus materials, and scoring materials (Including answer keys, rubrics, and plausible student responses matched to each of the score points on the rubrics) adhering to PARCC specifications described in Sections V.A.1.A.-V.A.1.C.

Response

Experience to Deliver Solid Item Development

The PARCC content development effort is a large and complex undertaking, requiring development of up to 6,800 items each year for ELA/literacy and mathematics. Pearson, ETS, and WestEd staff have experience working with PARCC to develop items/tasks and passages for the Phase 1 and 2 item development program. This experience will provide a solid foundation for future item development and will provide PARCC with a highly experienced team to continue to explore ways to introduce innovation into the item development process.

Recognizing that the amount of items/tasks/texts that need to be developed annually and with this work beginning in the summer of 2014, ETS, WestEd, and Pearson are ready to collaborate and begin work on the development of future items/tasks/texts. This collaboration will provide flexibility, capacity, and a deep knowledge base, enhancing the quality and efficiency of the content that is developed.

Pearson will provide the overall content development leadership and oversight for the program. Staff with PARCC content development experience will review materials prior to their presentation to committee review teams.

Additionally, the current leads on the current PARCC item development projects will provide leadership to maintain continuity between various phases of development. The following figure shows the proposed distribution of item development among the three organizations.

Content Area	**Responsible Party**
ELA/L Leadership	Pearson
ELA/L Grades 3-6	Pearson
ELA/L Grades 7 and 8	WestEd
ELA/L High School	ETS
Mathematics Leadership	Pearson
Math Grades 3-5	Pearson
Math Grades 6-8 and High School	ETS

This section will discuss the item development strategy meetings as the starting point for each cycle to plan for our success during the item development effort. Topics to be covered in the item development strategy meetings include:

- Item development targets for both ELA/L and Mathematics
- Item types that will need to be developed
- Technology-enhanced items
- Paper and pencil items

Item Development Strategy Meetings

Prior to the start of any item development cycle, a strategy session will be held. Pearson, ETS, and WestEd staff along with PARCC participants will meet to start each development cycle with a common understanding. Prior to the meeting Pearson and its subcontractors will conduct a review of the item banks for both ELA/L and Mathematics.

After the review of the bank, an item development plan will be developed for the upcoming development cycle. That plan will be presented at the strategy session and PARCC will then review and give input to the plan. Once approved, the item development plans, including specific targets by task type or evidence statement, will be the basis for that year's item development. Any proposed changes or modifications would be reviewed and approved jointly by PARCC and Pearson.

One item development strategy meeting is planned for each year, six weeks prior to the beginning of the development cycle. This meeting will need to happen soon after contract award in 2014 so we can stay on schedule for the required text and item reviews in the fall. Our assumption is that each strategy meeting is a daylong, virtual meeting.

Item Development Targets

ELA/Literacy Passages

We are committed to following CCSS and PARCC guidelines to find appropriate and engaging texts for all grade bands. An important consideration for our team as they engage in the search for stimuli is to be aware of associated costs. Our team has significant experience obtaining permissions for passages and multimedia. They have identified publishers and sites that can more affordably provide permissions to contain costs.

At a most basic level, our team will target texts that meet the set ranges made public by PARCC (as shown in the following figure).

PARCC's Set Ranges for Text			
Common Core Band	**The Lexile Framework**	**Reading Maturity**	**SourceRater**
2nd–3rd	420–820	3.53–6.13	0.36–5.62
4th–5th	740–1010	5.42–7.92	3.97–8.40
6th–8th	925–1185	7.04–9.57	5.85–10.87
9th–10th	1050–1335	8.41–10.81	8.41–12.26
11th–CCR	1185–1385	9.57–12.00	9.62–13.47

The Pearson ELA item development team is well versed in the CCSS and PARCC requirements for text complexity. CCSS champions a three-part model for measuring text complexity, which is comprised of qualitative, quantitative, and reader and task considerations. PARCC has implemented qualitative and quantitative measures for the text selection process.

As the team evaluates the qualitative measures of a text, they consider levels of meaning, purpose, structure, language conventionality, clarity, and knowledge demands using PARCC's Text Complexity Rubric. The ELA item development team has great familiarity with this rubric, and is comfortable and competent in its use.

For the quantitative measures, the team will use Lexile and Pearson's RMM (Reading Maturity Metric). Pearson will evaluate the impact of implementing other programs desired by

PARCC. For Reader and Task considerations, the team will assess factors such as student motivation, cognitive capabilities, level-appropriate knowledge, and alignment to task models when appropriate.

The five criteria established by PARCC to determine whether texts are worth reading will be used as guidelines while pursuing texts:

1. Texts are complex
2. Texts are diverse
3. Texts are authentic
4. Texts are paired effectively
5. Texts meet demands of bias and sensitivity guidelines

Appropriate passages for PARCC must also reflect the three key shifts represented by the CCSS that are at the heart of the ELA/Literacy standards. These shifts are:

1. Reading and writing grounded in evidence from the text
2. Regular practice with complex text and its academic vocabulary
3. Building knowledge through content-rich nonfiction and informational texts

Our ELA item development team is familiar with all of these requirements, and has demonstrated success selecting texts and multimedia that are suitable and engaging to students. The team also understands that all passages and/or multimedia must be appropriate for the intended Task Generation Models for which they will be used.

To contain costs, we will seek appropriate open source and public domain text for each grade band before pursuing permissioned texts. We also will apply the same philosophy in our pursuit of appropriate multimedia. We acknowledge that commissioned texts are not to be used on PARCC assessments. Permissions will be acquired for a minimum of three administrations (including one field test and two operational administrations).

Additional special considerations and limitations are to be considered when searching for appropriate ELA stimuli. For example, public domain text for students in the grades 3–5 band is generally more difficult to find than text for students at the upper grades. It is necessary to advise passage searchers to avoid several popular websites, particularly YouTube, due to permissions issues over multiple copyright holders for the same product. Perhaps most importantly, all texts and multimedia must be able to support items developed and aligned to CCSS.

Pearson and its subcontractors have costed appropriately for permissioned texts that will be newly developed for assessment administrations during this contract.Based on PARCC's instructions, we have included copyright permission costs that allow for the public release of two blueprints worth of ELA/literacy per year, assuming at least one third of the passages will be public domain. We also agree to procure permissions for passages chosen for the operational assessment that have permissions which were procured under a different PARCC contract, but will expire during the term of this contract. It is assumed at all passages from the other PARCC contract have a minimum of 3 year term. In the event the permission cost for a desired new or existing passage exceeds the tolerable costs limitations for a 3 year term, Pearson will consult with PARCC to confirm acceptance of a shorter permission term for the passage.

Copyright Clearance Center

Pearson is evaluating the feasibility of procuring a student assessment license from the Copyright Clearance Center (CCC) to meeting copyright permission requirements.

The student assessment license may offer an alternative to traditional, publisher-by-publisher, transactional licensing. Built around the concept of collective licensing, it includes a pre-authorized and growing repertoire of rights covering over 350,000 books, thousands of journals and magazines, and millions of news articles. The license also includes other media such as images and audio clips.

More than 40 publishers currently participate, including the National Geographic Society, The New York Times, Houghton Mifflin Harcourt, Carus Publishing Corporation, and the Hachette Book Group. Collectively, the student assessment license covers the use of hundreds of millions of authentic passages across every genre, grade, and reading level for possible selection.

All participating works are pre-authorized for assessments. The only required step in obtaining permissions is confirming that the desired passages are included in the student assessment license inventory of titles. This eliminates the advance administrative costs of pursuing individual permissions and the post-development administrative cost of maintaining those individual licenses. It also renders unnecessary the four month permission cycle mentioned at V.C.1 3.b, pg. 52, and eliminates the need for removal of items that occurs when individual permission are denied or delayed.

All titles in the student assessment license repertoire have a uniform set of reuse rights that include both print and electronic rights, the right to present the material for public comment, and the right to continue to use the materials for the full life cycle of an assessment passage, as long as annual license fees are paid. In the event Pearson, with PARCC's approval, determines that the student assessment license will be a viable solution, Pearson will pay the annual license fee as part of the base contract, during the term of the contract. In the event PARCC wishes to continue to use the licensed content after the term of Pearson's contract, PARCC may contract directly with CCC for an extension of the license term. As such, the use of the student assessment license may reduce the cost of tracking permissions, as well as risks associated with infringement or inadvertent misuse resulting from divergent terms and conditions attached to individual permissions, while at the same time increasing test security.

The license fee is an annual per student fee that covers the entire annual assessment cycle for English and Mathematics. Content may be added or removed across participating publishers and media throughout and across license terms without incurring additional permission fees.

The license fee scales to the number of students in the participating member states of the PARCC program. Each year, the licensee reports the actual number of students participating in the assessment program at specific grade levels. This count is multiplied by the agreed per student rate to determine the license fee. There is no need to over-license in anticipation of a possible expansion of the PARCC program. The Copyright Clearance Center receives reporting from the licensee on the actual usage and distributes the license fee in alignment with the actual use.

The license contemplates that licensees may change contractors over time, and the Copyright Clearance Center can, upon request, work with PARCC and/or its contractor to transfer the rights and permissions the current contractor acquires to any subsequent contractor or to PARCC directly with minimal paperwork, for the annual agreed upon license fee. In the event the CCC license is approved, Pearson will pay the license fee during the term of its contract with PARCC as part of the base contract price. PARCC would be responsible for the cost of maintaining the license after Pearson's contract ends.

The student assessment license offers the following benefits:

- **Extensive Content Pool from Leading Publishers.** Focus exclusively on the value of a given passage to assess student performance and adjust content as needed without incurring additional licensing fees or contending with potential delays.
- **Content Decisions not Informed by Permissions Costs.** The student assessment license provides one price for all permissions, and thus public domain and open educational resources materials only need be included if desired for pedagogical reasons.
- **Operational Efficiencies Through "Check-and-go-Permissions."** May save time and administrative costs otherwise spent obtaining copyright permissions on an individual basis.
- **Uniform Set of Reuse Rights.** The license provides the rights and flexibility you need to include copyrighted materials in your assessments and related preparatory materials.
- **Advance Knowledge of License Fees.** May make it easy for administrators to accurately budget and appropriate funds for copyright permissions ahead of time.
- **Increased Test Security.** Since there is no need to report content usage until well after an assessment has been administered—and even then the reporting is not at the individual text selection level—the chances of leaked test content is greatly reduced.

- **Migration from Paper to Digital.** Content under the license is covered for both print and digital formats. Migrating assessments from paper to digital does not require additional permission.
- **Multimedia Content.** The license includes images and other media that can replace materials commonly excluded by publishers when granting third-party permissions.

The student assessment license may provide substantial efficiencies in the contractor's acquisition of rights and in the rights holders' delivery of rights. The pricing for the license will capture these efficiencies and may lead to a lower average cost of acquisition for each copyrighted passage when compared with the full cost of acquiring permissions through the traditional publisher-by-publisher licensing method.

In addition to the direct efficiencies, by providing a uniform set of reuse rights and comparatively stable repertory of works, the student assessment license may reduce any on-going costs and complications of maintaining the PARCC assessment permissions over the medium and long term.

Pearson is also aware of additional special considerations and limitations that may be necessary to consider when searching for ELA stimuli based on RFP guidelines.

Perhaps most importantly, all texts and multimedia must be able to support items developed and aligned to CCSS. The Pearson ELA team understands that the delivery of appropriate passages is critical to the development of appropriate items, and that only strong text will supply a source for the "questions that beg to be asked."

Pearson is currently working with CCC to pilot using their repertory to search for passages based on criteria that we use for the PARCC program. The results of the pilot will be available by early spring. If the pilot is successful, Pearson will pursue obtaining a student assessment license with CCC for use by our contractors and our content staff. Although we expect the license to be cost neutral compared with in-house permissions (primarily because of the anticipated need to regularly obtain at least some permissions by going outside of the license), in most cases the use of the license should significantly reduce the time needed for obtaining permissions and provide more flexible conditions for using permissions.

ELA/Literacy Items

No items for PARCC are developed without a specific purpose, as the expectation is that all items clearly align to the assessment's purpose from origination; these items are not aligned after-the-fact for any purpose.

Our ELA item writers have formal training in their areas of expertise as well as classroom and large-scale assessment experience in most cases. Strong items rely on our ability to select stimuli that reflect appropriate text complexity. Without the appropriate foundation, our items will not align properly to CCSS or PARCC's evidence statements and Task Generation Models.

Successful ELA items allow students to demonstrate that they are successful independent readers. As per Appendix A of CCSS, students must be able to identify, evaluate, and use evidence to support or challenge a thesis. They must also be able to consider and incorporate counterarguments into writing. Items developed must meet the demands of PARCC and allow students to:

- Read and comprehend sophisticated texts independently
- Make, defend, or dismantle an argument
- Interact with unfamiliar vocabulary
- Use evidence effectively
- Conduct research for a variety of purposes

Our team has experience and ability to meet the directive to develop items that allow students to demonstrate their ability to locate and deploy evidence effectively, as these skills are hallmarks of strong readers and writers. If students are not allowed to show these abilities, the assessment will fail to determine if students are indeed college and career ready.

All items developed for PARCC ELA must be text-dependent. The items cannot be low-level, literal, or recall questions, and they cannot be focused on comprehension strategies. Ideally, these questions should involve analysis, synthesis, and evaluation; students must be required to use evidence from the text to correctly answer questions that are worthy of inclusion on the PARCC assessments.

Math Items

In mathematics, planning for development requires application of the shifts outlined in the Common Core State Standards for Mathematics (CCSSM) of focus, coherence, and rigor.

Focus. The bank of tasks should reflect the major work of a grade or course. We recognize, however, that operational development must be informed by field-testing development and results. Pearson and ETS will work with PARCC to determine development targets based on PARCC's requirements and the bank resulting from field-testing. By focusing where the standards focus, we can develop an assessment that meets Claim A and reflects the important work being done in the classroom by teachers and students.

Coherence. An assessment system should allow students to connect mathematical concepts together. This can take the form of connecting content within that course or grade; however, it can also require students to connect content to key concepts from previous grades or courses. PARCC's blueprints already reflect this type of coherence. Our team will use the blueprints to inform development. Tasks that align to a cluster heading or a domain/course will be developed to reflect the coherence within that grade/course and will also help meet Claim B.

We will also be mindful of the appropriate use of prior knowledge, particularly as it pertains to the Performance-bases Assessment and the development of tasks aligned to Claims C and D. These tasks often require content from a previous grade or course be used as the backdrop for students to demonstrate their ability to reason or model. Our staff welcomes the opportunity to weave in content from previous grades or courses. Developing tasks at a broader alignment also reflects what we hope to see in the classroom as the CCSSM continues to be implemented.

Rigor. The CCSSM defines rigor as a balance of procedural skill and fluency, conceptual understanding, and application. This balance should also be evident in a bank of tasks designed to assess the CCSSM. We have experience successfully developing tasks in all three areas and recognize that procedural skill and fluency, while only reporting out for Claim E in grades 3–6, is important in all grades and looks different as a student moves from grade 3 through high school.

Similarly, the types of applications and concepts students will be asked to demonstrate and use change as students' mathematical understanding increases. With PARCC's input, we can develop tasks that meet the intended rigor of the CCSSM while filling the needs in the bank.

The mathematical practices should be a major consideration when planning development of a CCSSM assessment. All claims require connecting to the practices and our team is experienced in making these connections. We also recognize that the practices cannot be separated from the development plan as an afterthought or metadata. We will work with PARCC and use the tools provided by the consortium to develop an assessment where the practices are a major part of the training and task-writing process.

Item Types

ELA/Literacy

When developing specific ELA item types, every item must test reading claims RI/RH/RST/RL 1.1 and/or 1.2 of the CCSS. In addition, vocabulary items are required to test Tier II words at each grade level, not domain-specific terms that may appear in passages that must fulfill requirements for history/social studies and science/technology topics. In addition, every item must test additional reading claims once the first requirement is met. It is desirable for individual items to test multiple standards whenever possible.

Item types are prescribed within each task type, and our item development team must adhere to specific task requirements. The complexity of items is tracked by the team, and the team will always bear in mind that the most desirable way to create items of high text complexity is to have passages of high text complexity.

A key requirement is for a range of difficulty and rigor to be reflected among all item types within each item set at each grade. As items are developed, the item development team must target the three key shifts at the heart of the ELA standards:

1. Reading and writing grounded in evidence from text
2. Regular practice with complex text and its academic vocabulary
3. Building knowledge through content-rich nonfiction and informational texts

With these shifts in mind along with appropriate use of the evidence statements, there are three item types that the ELA team must develop for PARCC:

1. Evidence-based selected-response items
2. Technology-enhanced constructed-response items
3. Prose constructed-response items

Evidence-based selected-response items and technology-enhanced constructed-response items are designed to measure reading, while prose constructed-response items are designed to measure reading and writing (except in cases when a narrative writing response is being elicited).

An evidence-based selected-response item may be used for Performance-Based Assessment and End-of-Year Assessment. Evidence-based selected-response items need to allow machine scoring, and these items usually consist of two parts that work together as a single item. A student working through an evidence-based selected-response item will first see Part A of the item, which will pose a question and present at least four options. These items may be single or multiple-select, with directions instructing students as to how many options they must select.

Once the student has answered the Part A portion, they proceed to Part B. In Part B, the student must identify the evidence that explains how they arrived at their answer in Part A. Again, this part of the item may be single or multiple-select.

Due to the relationship between Part A and Part B of an evidence-based selected-response item, the possible evidence in Part B must parallel the plausible options in the first part, all of which must be based on incorrect text-based inferences or incorrect conclusions from not reading carefully. The number of answer choices required in parts A and B is mandated based on the number of correct responses within the item part (for example, if there are two correct responses, six answer choices should be provided).

A technology-enhanced constructed-response item may be used for Performance-Based Assessment and End-of-Year Assessments administered online. A technology-enhanced constructed-response item should allow for machine scoring, and relies upon functionalities such as drag and drop, hot spots, text flagging, text extraction, hovering, annotation, and so forth. The technology must be important to the assessment of the skill; PARCC is not looking simply for items with technology added. If an item can be asked in a multiple-choice format, it should most likely be presented as an evidence-based selected-response item rather than as a technology-enhanced constructed-response item. The technology must serve to integrate ideas from the text and allow the student to show conclusions they make while reading. These items are usually one-part items, with the evidence embedded in the technology-enhanced portion of the item.

The directions for technology-enhanced constructed-response items must make it clear to students what they are expected to do, as this item type will take a while to become familiar to students. Device-neutral language must be used within technology-enhanced constructed-response items, as PARCC assessments may be delivered across a variety of devices. Therefore, terms like “click” should be abandoned in favor of neutral language like “select.”

A prose constructed-response item may only be used on the Performance-Based Assessments, as this is not a machine-scorable item type. An effective prose constructed-response item must make it clear to the student why they are writing (establish a clear purpose) and to whom they are writing (specify the audience). The topic must be clear, and a successful response demands that the student use evidence and/or details from the associated stimuli.

For prose constructed-response items to elicit meaningful and thoughtful responses from students, topics must be engaging, relevant, and clear. Students will be scored based on their ability to provide evidence for the sub-claims for Written Expression, Conventions and Knowledge of Language, and at least one Reading standard that falls within the major claim of Reading Complex Text.

Math

Pearson and ETS will develop mathematics tasks that address the Common Core State Standards and meet the expectations presented in the course level blueprints and test specifications. Pearson has full knowledge that each grade/course has a set of blueprints and set task types for the individual test design. Through the use of the blueprints and form specifications, we will adhere to the content limitations of each evidence statement, along with assuring that the tasks assess the cluster heading of the CCSS. We will address one or more of the mathematical practices throughout each Task Type. The various task types are explained in the proceeding paragraphs.

For Type I tasks, we will develop tasks that include the assessment of major and supporting content along with fluency, by using Claims A, B, and E. Pearson’s use of multiple choice and multiple select will be used for this item type, however, technology enhanced items will be the forefront for development. Technology-enhanced items will include the accurate usage of technology so the evidence statement being assessed is through the student’s knowledge of the content, rather than the opportunity for choosing a correct answer based upon elimination.

We continue to explore the possibility of rules-based scoring dependent prompts while still allowing students to receive partial points for the individual prompts of the tasks. Type I tasks include more than 1 point value at times, therefore, the use of single and multiple prompts will be developed.

For multiple prompt tasks, the prompts will correlate with each other, but will not include dependency. This allows for the student the opportunity to receive partial points for an individual prompt. Additionally, multiple prompt tasks may include both a technology enhanced prompt along with a multiple choice, multiple select, or constructed response prompt.

For many Type II tasks, we will develop tasks that include the assessment of Claim C to identify the students' knowledge and understanding for reasoning. These tasks will ask students to construct viable arguments, critique the reasoning presented in an item, attend to precision when making mathematical statements, and lastly, incorporate content from Claim A or rarely Claim B. Type II tasks can involve both machine and human scored prompts allowing the student to answer one prompt through identification and then allowing them to explain or show their work for the reasoning portion of the task.

All Type II tasks will involve at least 50 percent of the allotted points, either a 3-point or 4-point value, to the reasoning portion of the evidence statement using one or more of the Common Core State Standards. Within a multiple-prompt task, special attention will be paid to avoid cuing amongst each other.

For Type III tasks, we will develop tasks that include the assessment of Claim D to identify the students' knowledge and understanding for modeling. These tasks will ask students to solve real-world problems with a degree of difficulty appropriate to the grade/course; apply knowledge and skills articulated in the standards for the current grade/course (and possibly previous grades/courses); and lastly, incorporate content from Claim A and other practices. Type III tasks can also involve both machine and human scored prompts similar to Type II tasks.

All Type III tasks will involve at least 50 percent of the allotted points, either a 3-point or 6-point value, to the modeling portion of the evidence statement using one or more of the Common Core State Standards for mathematical content.

Throughout the development process of various task types, our team will continuously strive to introduce creativity and new technology enhanced items to enrich the quality of the PARCC assessment. Our content team regularly collaborates with other professionals to discuss ways to add innovation to the measurement of the Common Core State Standards.

Technology-Enhanced Items

PARCC's commitment to exploring and expanding the role of technology-enhanced items in their assessment is obvious through their current work on the assessment and the goals of use of these items types as described in the proposal for continued item development.

Our team is also striving to embrace the use of technology-enhanced items and wants to work with PARCC in the continued development of new assessment strategies that allow a student to show what they know beyond the limitations of a paper-pencil test. It is important to work as a team to make sure that all new technologies introduced on the PARCC assessment bring true value and are not just extraneous in nature.

Experienced with Technology-Enhanced Item Development

We are familiar with the five categories of technology-enhanced item development that is shown in Attachment H. This attachment documents some of the item types that we have currently developed in conjunction with Phase 1 and we will continue to work on making technology-enhanced items a significant part of the PARCC test by increasing the use of technology-enhanced items systematically over the life of the contract. The proportion of technology-enhanced items will be mutually agreed upon at the beginning of each development cycle and will not exceed 50 percent.

The types of technology-enhanced items that are appropriate for the content change as students move through their education; however, we are always working to find new ways to creatively but effectively use our current functionality. We do not want to rule out a particular functionality.

The PARCC Spring 2014 Field Test will include equation editor items that allow students to enter open-ended math responses consisting of numbers, a single mathematical expression, or an equation that includes mathematical operators and functions not found on a standard keyboard. These items are scored automatically by Pearson's MathQuery scoring engine using a rule-based approach.

For a given equation editor item, the scoring rubrics specified by the item writer are encoded into MathQuery using domain modeling techniques similar to those found in intelligent tutoring systems. When the MathQuery engine receives a student response, it looks for characteristics of the response that match the scoring criteria, and assigns the appropriate score. This analysis goes beyond simple matching against a list of known answers to actually reasoning about the response.

Many automated math scoring systems evaluate only whether a student's response is equal to a known correct answer. When equality is the only rubric, the information that can be identified in a student response is extremely limited, so the assessment tends to focus almost exclusively on computation skills.

The Common Core State Standards seek to balance computational proficiency with mathematical understanding, as reflected in the “Standards for Math Practice.” MathQuery uses equality testing when appropriate, but it can also identify complex patterns, key relationships, problem-solving strategies, and other factors that can be used to assess critical thinking skills. Using this approach, MathQuery can determine whether a student understands a core concept—what it means to write the equation of a line in standard form, for example—independently of whether the student can execute the problem correctly. We will continue to develop equation editor, technology enhanced items as part of the PARCC operational delivery.

Working with PARCC on New Item Development

New development of technology-enhanced items can be challenging as new item types run the realm of development from new item interactions to new scoring schemas to untested item concepts. We will work with PARCC to introduce these types in a systematic manner that allows proper documentation and preparation to verify the success of the item type. At the heart of every technology change should be the concept that this item type is bringing a documented value to the way in which a student can respond.

Innovative ideas for technology-enhanced items can be generated in various ways: content specialists who are working with the evidence statements realize an item can be written in a new way to meet that standard, comments and ideas gathered at committee meetings where PARCC members are discussing item constraints or limitations, our internal research group that focuses on new item type development proposes a new idea.

Our subcontractors also will bring forward ideas for innovation. These ideas need to be documented and evaluated by the PARCC item development group. These ideas will be periodically reviewed and evaluated before a new item type is brought to the PARCC Partnership group for approval.

Factors used in evaluating a potential technology-enhanced item functionality include the following:

1. Does the item fit within PARCC’s vision for their assessment?
2. Does the functionality allow for the authentic assessment of the evidence statement in a new way?
3. Is this new item type cost-effective for the subject matter and breadth of content that it can be used for?
4. Are there any limitations in technology and/or scoring schemas that cannot be overcome in a timely manner?
5. Does the item functionality fit within PARCC’s technical specifications?
6. Does the item functionality fit within PARCC’s need to balance accessibility considerations?

If these and other factors are evaluated and it is deemed appropriate, the new technology-enhanced functionality item types will be brought to the PARCC Partnership group for their evaluation and approval before moving forward with development. New technology-enhanced items can be difficult to develop at times and unforeseen issues can appear as item development proceeds.

By working together in this organized process, PARCC and the Pearson item development group can deal with these challenges to make the best decisions on new technologies and item functionalities.

Embracing the idea of a structured approach to the introduction of technology-enhanced items to the overall PARCC assessment with the ultimate desire to be 50 percent technology-enhanced items, Pearson outlines the following item development plan. The following figure shows our ability to maintain the specific annual technology-enhanced item targets while incrementally increasing the number of technology-enhanced items each year until the targeted percentage is reached.

As part of the discussions around development targets, Pearson will work with PARCC to determine mutually agreeable category percentages for technology-enhanced functionality types shown in Table V.A.1.C.4 of the RFP, while also considering the impact of increasing the overall item pool with technology-enhanced items. Over the course of the contract, we anticipate that no more than 6 new functionality types will be added to the table.
The following figures show Year 1 counts for technology-enhanced items in both Math and ELA/L to be available for field testing. These are the minimum counts that will be available in future years of the contract. Pearson and its subcontractors will work with the Partnership to adjust totals upward to meet PARCC's long-term desires around technology-enhanced items.

Math Development Totals for Year 1		**ELA Development Totals for Year 1**	
Grade 3	44	Grade 3	49
Grade 4	41	Grade 4	49
Grade 5	38	Grade 5	49
Grade 6	40	Grade 6	49
Grade 7	40	Grade 7	50
Grade 8	40	Grade 8	50
Algebra I	62	Grade 9	71
Geometry	61	Grade 10	71
Algebra II	44	Grade 11	50
Mathematics I	41	**Total**	**488**
Mathematics II	41		
Mathematics III	43		
Total	**535**		

Reasons to use technology-enhanced items in assessment are many, but the most obvious is that it is in keeping with the sense of autonomy today's learners have embraced with the use of technology in their everyday lives. Technology-enhanced items need to test standards in ways that evidence-based selected-response items cannot; in other words, the use of technology should be integral to the content of the item rather than a novelty.

Testing Student Ability

If a technology-enhanced item can be asked as an evidence-based selected-response item, it should be. Technology-enhanced items seem most appropriate for testing students' abilities to sequence events, follow or arrange steps in processes, produce summaries, annotate texts, and other similar skills.

Although some of the available technology-enhanced item types are enhanced multiple-choice items, TestNav has introduced templates that parallel how students learn literature, writing, and English in a classroom. Educators teach students to think and explore in a text rich environment, and a PARCC assessment should allow students to respond to questions in the same way.

To assess higher-level cognitive abilities, we need new ways to allow for the presentation of responses. An assessment rich with technology-enhanced items aligns with how students learn, and, ideally, becomes less artificial and more like the world in which a student meets and examines information.

The idea behind an ELA technology-enhanced item is that it can encompass the same flexibility and adaptability reflected in the 21st century classroom. Students learn to classify, categorize, and connect ideas and graphics in the real classroom, and a technology-enhanced item allows a student to mirror their learning more realistically. They can match text or graphic representations from passages and rearrange words, phrases, sentences, and paragraphs. Learners have the opportunity to extract or select text from a passage and reorganize or highlight it to demonstrate learning.

Again, we do not want to use technology enhanced items just because we can; we want to use technology-enhanced items because they enhance a student's ability to achieve success due to the parallel of skills they have and are exposed to in the classroom as well as in their daily interactions.

For example, younger children have gaming skills that include point and click, drag and drop (more by touch screen than mouse these days), and picture matching. We attempt to remain mindful of appropriate grade-level ELA skills as we develop technology-enhanced items, and then incorporate students' skills into the development of our technology-enhanced items.

We have included a variety of samples that showcase the range of innovation in technology-enhanced items that we are currently developing. We continue to look for ways to increase our item development capabilities beyond what is shown here. Below are two sample tasks that push innovation.

Sample Task 1

As described below, the innovations of the proposed task are threefold.

1. Create a New Structure for a Task

The research task will create two research topic tracks that essentially create two test forms that share the same anchor passage and a set of items. After the students read the anchor passage and answer traditional evidence-based selected-response and technology-enhanced constructed-response items, the students choose one of two topics, based on the anchor passage, to research further for a scenario-based research paper's thesis.

Students on both research tracks encounter similar item formats, though the content they encounter is specific to their chosen research track. The task scaffolds the students to a culminating prose constructed-response item.

2. Target Common Core State Standards not yet Fully Assessed

The research task targets Reading (RI; RH), Language, and Writing standards. However, CCSS not fully addressed by current Performance-Based Assessment (PBA) Research Simulation Task Generation Models, namely Writing standards that target research of sources and evaluation of sources, are the innovative focus of the research task.

3. Use Current Technology and Functionality in New Ways

The research task uses current PARCC assessment technology and functionality in ways not yet accomplished. For example, students use hot text functionality on images of authentic newspaper front pages and select texts in a simulated Internet search.

For the purposes of the proposal, we limited the innovative task to researching, evaluating, and selecting relevant materials for a research paper. The task, and its innovation, is expandable to include a step whereby students select graphics relevant to their topic, such as newspaper headlines, photographs, maps, and cartoons, and take notes on a digital text related to their research, and place all materials in a virtual folder.

Students then use these materials to write a more elaborate response to a follow-up technology-enhanced constructed-response item item or create PowerPoint-like slides for a class presentation. In this manner, PARCC would be able to more truly simulate a classroom-based research project than is currently possible.

In this sample task, students are asked to read a passage about the start of the Spanish-American War and certain events that may have precipitated the United States' involvement. Specifically, the influence of William Randolph Hearst and yellow journalism and the sinking of the USS *Maine* are discussed in the anchor passage.

After reading the passage, students are administered evidence-based selected-response and technology-enhanced constructed-response items that focus on vocabulary, comprehension, and an understanding of the structure of the text. Students are then administered two additional technology-enhanced constructed-response items that require them to apply what they have learned about yellow journalism. In the first technology-enhanced constructed-response item, students categorize newspaper headlines as being either examples of yellow journalism or examples of traditional journalism. In the second technology-enhanced constructed-response item, students identify those words and phrases that make a particular headline an example of yellow journalism.

At this point, students are allowed to choose which of two areas of concentration, or tracks, they would like to research further: William Randolph Hearst and his possible influence on the United States' decision to declare war, or the effect that the sinking of the USS *Maine* had on the outbreak of the war. Depending on which track the students select, students are administered an item that focuses their attention on evaluating and selecting appropriate and relevant texts for their research.

Presented to appear like a typical Google or Yahoo! search result, students select two texts they feel are most relevant to their research. Thus, this source selection is in actuality a multiple-select multiple-choice item. A subsequent splash page informs the students of which texts are truly most relevant, so that students who selected a text that is only somewhat related to the subject (e.g., the Hearst Castle) are still asked to read only the most pertinent texts.

Students then read the two texts on their chosen track and answer evidence-based selected-response and technology-enhanced constructed-response item items, just as they did with the anchor text. The texts for each track have been carefully evaluated in terms of their word length and their Lexile© score so that both tracks are of similar complexity (as shown in the following figure). Finally, all students respond to a prose constructed-response item that asks them to write an essay in which they evaluate the two texts and explain why each source is relevant to the topic and is reliable.

Word Lengths and Lexile© Score of Task Passages				
Passage Title	**Author**	**Source**	**Word Length**	**Lexile Score**
The Newspaper War	Jason Skog (main text) Samuel Willard Crompton	*Yellow Journalism* (We the People) Sidebar source: *The Sinking of the USS* Maine*: Declaring War Against Spain*	910	1050L

Word Lengths and Lexile© Score of Task Passages				
Passage Title	**Author**	**Source**	**Word Length**	**Lexile Score**
	(sidebar)			
Track 1				
The Crucible of Empire: William Randolph Hearst	PBS	http://www.pbs.org/crucible/frames/_journalism.html	495	1170L
W. Joseph Campbell Letter to the Washington Post	W. Joseph Campbell	http://academic2.american.edu/~wjc/documents/remington_hearst_wjc_letter.pdf	380	1180L
Total Word Length and Average Lexile©			**875**	**1175L**
Track 2				
The Sinking of the USS Maine	John F. Wukovits	*The Spanish-American War*	483	1180L
Explosion Aboard the USS Maine	The Learning Network (*New York Times*)	http://learning.blogs.nytimes.com/2012/02/15/feb-15-1898-u-s-battleship-maine-explodes-in-havana-harbor/?_r=1	426	1030L
Total Word Length and Average Lexile©			**909**	**1105L**

Sample Task 2

The Performance Task provided, “A Comparison and Contrast of Fictional and Nonfictional Accounts of the Civil War,” is a sample of a grade 7 task that pushes the boundaries of assessment. It is based on four authentic texts and culminates in a writing prompt that is a direct and robust measure of a key Common Core Standard—Reading Literature, Standard 9. The innovative nature of the task lies in five features:

1. The crafting of an engaging, authentic scenario and “Questions to Consider,” both of which help to create an assessment task that mirrors classroom instruction;
2. The selection of four complex, thematically connected texts, one non-fiction and three fiction, to provide a solid foundation for a difficult-to-assess standard;

3. The writing prompt, which assesses both reading and writing, will elicit thoughtful responses to a question that asks students to compare and contrast the depiction of characters and events, fictional and actual, in the same historical period;
4. A short constructed response item (item 2 following the Charley Skedaddle text), which may serve as a model for a constructed-response item type with the potential for automated scoring;
5. The provision of a writer's toolkit, including resources for drafting and outlining the response and a writer's checklist; also included is a mockup of a tool to auto-populate a Works Cited page with bibliographic information about each text.

The first text the students read is a passage from *Sarah Morgan: The Civil War Diary of a Southern Woman.* As described in a purpose-setting statement, this anchor text is a diary of a young woman who lived and wrote from her home in Louisiana during the Civil War. With their wealth of detail about people and events, the nonfiction diary entries provide the historical account of a time period that will become the basis for the writing task. The "Questions to Consider" remind students to think about what Sarah does and says and how she feels about the war, all of which they can later cite as evidence for their responses to the central focus of the task and of Reading Literature standard 9 at Grade 7: the comparison and contrast of fictional and nonfictional accounts of the same period as a way of understanding history as it is portrayed by writers of fiction. Following the reading of the text, students are also given a selected-response question. Since the text is Informational, the item aligns to standard RI 6, which asks about an author's point-of-view or purpose in a text.

The second text, *Shades of Gray,* provides a fictional perspective on the war. In its focus on characters and their efforts to grapple with and understand the war, the passage is a rich source of ideas on the question of how authors of fiction use or alter history. In particular, the "Question to Consider" about the novel's title is meant to suggest to students that even though the story is told from the perspective of the South, the characters' views on the war vary. Once the students have read the text, they will be required to respond to two items, both technology enhanced. The first, a two-part item, asks about both the characters' opinions about the war and a technique used by the author to develop and reveal the characters' points-of-view. It is closely aligned to RL 6, which asks not only about characters' or narrators' points-of-view but also about how an author develops and contrasts points-of-view in a text. The second item about this text aligns to RL 3, which requires an analysis of how particular elements of a story or drama interact. In this item, the students must understand how the text's dialogue and the events in the text interact and shape the characters in ways that are distinguishable by their attitudes.

The third text, *Charley Skedaddle*, is another fictional account of the war. Told from the perspective of characters living in the North, it contrasts with an otherwise similar fictional account of the war, *Shades of Gray*. Through its rich dialogue, mainly expressed in the dialect of the time period and the place (New York City's Bowery neighborhood), the passage focuses on the character of Charley and his reasons (asked about in the "Questions to Consider") for wanting to fight in the war. It is followed by two items that are both technology enhanced. The first elicits an objective summary of the text in a machine-scorable item and aligns closely to the standard, RL 2, which requires an objective account of events; here, non-essential sentences must be eliminated from the provided summary and the correct reasons for their omission must be selected. The second item aligns to standard RL 6 and asks about the main character's attitude toward Confederate soldiers. It is particularly innovative in requiring a short constructed response. This item could be considered as a model for an item type that is yet unexplored in PARCC item development—a short constructed-response item that would be amenable to automated scoring.

The fourth text, *Across Five Aprils*, is another fictional account of the war. Like Charley Skedaddle, it uses rich dialogue that is expressed in the dialect of the time period and place; however, unlike the other two fiction passages, it represents the perspectives of both the North and the South. Students are asked to consider these perspectives through the "Questions to Consider," which ask about their differing views on the war, the decision of one character to fight in the war, and the main character's feelings about this decision. It is followed by two items. The first, a technology-enhanced item, asks students to identify the mood that the author creates by using imagery in the text. The second item, a multi-part selected-response item, focuses on the characters and their conflicts and aligns to RL 6.

The task culminates in a writing prompt that requires students to combine their reasoning, information, and evidence from all four sources. Using the nonfiction text as the anchor, the prompt poses three questions:

- How would Sarah Morgan feel about the characters and events in each of the three fictional texts? With which characters would she agree?
- If Sarah Morgan could talk to the characters of the three fictional texts, what might she tell them about her own experiences? How might she compare their fictional experiences to her own actual experiences?
- How did the fictional and nonfictional accounts of the Civil War change your understanding of the events?

In considering and responding to these questions, the students are directed to write a comparison and contrast essay that will assess their performance on standard RL 9 for grade 7. Through their reading, responses to selected-response and technology-enhanced items, a short constructed-response, and this extended constructed-response, the students will "*compare and contrast a fictional portrayal of a time, place, or character and a historical account of the same period as a means of understanding how authors of fiction use or alter history*."

As the students write their essay, they are able to use several tools that help them in drafting a response, citing their sources, and reviewing/revising their work. In particular, the Writer's Checklist helps to remind students of the features of a well-written essay, including the kind of thoughtful analysis and evidence needed to produce an excellent response. Following the texts, items, writing prompt, and toolkit, the task includes the keys for the selected-response and technology-enhanced items.

The following table displays the information about each of the texts, including length and quantitative readability. All of the fiction texts are appropriately placed at the Grade 7 reading level and include glossary terms for words that may be unfamiliar. The nonfiction text has higher readability scores, as measured by the ETS tool, TextEvaluator (formerly SourceRater), and by Flesch-Kincaid, but its diary format, high-interest content, and conversational style should make it accessible to Grade 7 students. As with the fiction texts, glossary terms are included to provide definitions of words that are unfamiliar or domain-specific.

A Comparison and Contrast of Fictional and Nonfictional Accounts of the Civil War					
Passage Title	**Author**	**Source**	**Genre**	**Word Length**	**Text Evaluator Score**
Sarah Morgan: the Civil War Diary of a Southern Woman	East, Charles	New York: Touchstone, 1992	Nonfiction (Informational)	1283	11.3
Shades of Gray	Reeder, Carolyn	New York: Macmillan Publishing Company, 1989	Fiction (Literary)	771	7.2
Charley Skedaddle	Beatty, Patricia	New York: Troll Communications, 1987	Fiction (Literary)	723	6.8
Across Five Aprils	Hunt, Irene	New York: Berkley Books, 1965	Fiction (Literary)	1969	8.2

Sample Items Redacted

Paper-Pencil Items

ELA/Literacy

Technology-enhanced items cannot be used on Pencil and Paper test forms. The item development counts include additional evidence-based selected-response items that we plan to use on the paper forms. No additional development has been planned.

Math

PARCC has identified the number of blueprints of items that are to be assessed on pencil and paper test forms. We acknowledge the need for paper-pencil items for those students who do not have the accessibility of technology or need such accommodations. We will guide the development of paper-pencil tasks by verifying that 50 percent of each Evidence Statement will be assessable on a paper-pencil format.

The development process for paper-pencil tasks begins with looking at the technology-enhanced items that need a replacement. We will first look at the construct of the original item and then develop the paper-pencil item to involve the same construct, while also taking into consideration the cognitive complexity.

At times, the same construct cannot be maintained between the technology-enhanced item and paper-pencil replacement, if that is the case we will make alternative decisions. First, we will try to find another Evidence Statement within the same probability cluster with a similar construct. If this is still not possible, we will develop a different construct, but using the same Evidence Statement.

The overall goal of a paper-pencil item is to not give advantage or disadvantage to those students who are taking the assessment online. We try to confirm that the complexity and mental synthesis is maintained between the technology-enhanced item and paper-pencil replacement.

Pearson acknowledges that paper-pencil items must go through the committee review process as usual and there are times when the technology-enhanced prompt is accepted, but the paper-pencil prompt is not or vice versa.

Requirement

V.A.1.D. Accessibility and Fairness

1. Strategy and Capacity for Developing Items Meeting Bias and Sensitivity Guidelines

Response Requirements for Section V.A.1.D.1

a) Offerors must provide evidence of their capacity and prior experience in producing items/tasks, stimuli, passages, performance tasks, online tools, and graphics in English language arts/literacy and mathematics that are sensitive and free of bias.

b) Offerors must describe their strategy for training item developers to use the Bias and Sensitivity Guidelines.

c) Offerors must describe their strategy for ensuring item developers will use the Bias and Sensitivity Guidelines.

d) Offerors must provide their strategy for training item developers to adhere to the linguistic complexity rubric.

Response

Developing Items According to Bias and Sensitivity Guidelines

According to the National Research Council (1999), "fairness, like validity, cannot be properly addressed as an afterthought once the test has been developed, administered, and used. It must be confronted throughout the interconnected phases of the testing process, from test design and development to administration, scoring, interpretation, and use" (p. 81).

Providing fairness, sensitivity, and freedom from bias in educational achievement assessments is a matter of informing and sensitizing item and task developers to guidelines for avoiding sensitive topics and other sources of bias for the students who will participate in PARCC assessments, the people who teach them, and their families and communities. It also requires systematic review of all stimulus materials, items, and tasks by experts in bias and sensitivity to verify that the guidelines have been followed successfully.

The PARCC Bias and Sensitivity Guidelines and our established Universal Design guidelines will converge to form a comprehensive set of guidelines and supporting tools, such as checklists, for purposes of training item and task authors and reviewers who will develop and review PARCC items and tasks.

Our Development Process

Bias and sensitivity are standard focus in our process for developing items, tasks, stimuli, passages, performance tasks, online tools, and graphics in English language arts/literacy and mathematics materials for assessments:

- Our team is committed to developing materials that are free from bias and sensitivity as well as accessible from inception. We provide extensive training for item and task writers.
- We employ experienced and well-trained internal specialists, including experts who review all items, tasks, stimuli, passages, performance tasks, online tools, and graphics in English language arts/literacy and mathematics materials for bias and sensitivity, as a standard step in our internal review process.

Pearson has successfully delivered PARCC items, tasks, and stimulus materials that are free of bias and sensitivity issues. During Phase I of item development our acceptance rate for passages is over 99 percent and over 98 percent for items. Our acceptance percentage across all large-scale state assessment programs since 2009 for which we provide high-stakes item development services has been greater than 95 percent.

All item developers for the Pearson collaborative, both writers and content specialists have been provided training utilizing bias and sensitivity guidelines. Our standard training has been evaluated with respect to the PARCC Bias and Sensitivity Guidelines. Some points related to assessing the construct have been highlighted. The Common Core State Standards for ELA/Literacy require source materials that fit into a gray area in some state programs, so additional discussion of what is and is not appropriate and required by the construct is emphasized in our training.

Items and passages are reviewed internally by experts trained in the bias and sensitivity awareness and the principles of Universal Design. Suggestions for edits are provided to content specialists to evaluate with respect to the construct being assessed and experience with PARCC's application of the bias and sensitivity guidelines. Feedback is provided to item writers along with follow-up training when necessary.

The linguistic complexity of an item can affect student performance, particularly those students where English is not their primary language. PARCC has implemented a process to evaluate the linguistic complexity of text through the use of a rubric and by assigning committee members responsibility for a linguistic review. Three areas are evaluated on the language complexity rubric: text density, language structure, and vocabulary. Pearson has addressed some of the pieces such as limiting the use of passive voice and using shorter sentences with fewer clauses, when appropriate for the cognitive demand and necessary for the construct or defensibility of an item.

Pearson will review the linguistic complexity rubric and incorporate it into our item review process. There is a tension between the text complexity and vocabulary requirements for the ELA passages and the linguistic complexity. Pearson will work with PARCC to determine the expectations for the use of rubric and the interplay between it and the text complexity rubric. Steve Ferrara has a research interest in cognitive complexity and linguistics. He will work with us to develop training materials based on information provided by PARCC for our internal content team and item developers.

Requirement

Response Requirements for Section V.A.1.D.2.

a) Offerors shall provide a 6-8 page statement in their response/reply which describes their plan for meeting the requirements listed in the section above.

b) The response will provide evidence of the Offeror's capacity and prior experience in using Accessibility Guidelines to produce accessible items/tasks, stimuli, passages, performance tasks, online tools and graphics in English language arts/literacy and mathematics.

c) Offerors must describe how item writers will consider and be trained to incorporate Accessibility Guidelines and basic Universal Design principles, including but not limited to the following:

    i. The item or task measures what it intends to measure.

ii. The item or task respects the diversity of the assessment population and allows the full range of eligible students to respond to the item/stimulus.

iii. Decisions will be made to ensure that items and tasks measure what they are intended to measure for EL students with different levels of English proficiency and/or first language proficiency.

iv. All supports have been considered that may increase access while preserving the targeted construct or a range of students with disabilities.

v. Multiple means of presentation, expression, and engagement have been considered with regard to individual items/tasks for both SWD and Els.

vi. The item or task material uses a clear and accessible text format.

vii. The item or task material uses clear and accessible visual elements (when essential to the item).

viii. The item or task material uses text appropriate for the intended grade level.

ix. Changes to the format of an item will be considered to preserve the item/task meaning or difficulty.

d) Offerors must describe their strategies for ensuring that item developers will use the Accessibility Guidelines.

Response

Meeting PARCC Requirements for Accessible Assessments

We share PARCC's desire to make the assessments accessible to as many students as possible, including students with disabilities and English language learners. PARCC plans to use an inclusive assessment system incorporating a combination of accessibility features and accommodations to support this goal. Features that are enabled by the delivery platform, some with the support of item development, have been classified into three categories:

- Accessibility Features and Accommodations
- Presentation Accommodations
- Response Accommodations

This set of features will enable PARCC to use content developed to the principles of Universal Design and deliver assessments that are fair to all students and that allow valid inferences about what all students know and can do. Our team is committed to developing assessments accessible to the widest population, while maintaining the constructs being assessed.

Applying the Principles of Universal Design

Pearson has applied the principles of Universal Design, as articulated in materials developed by the National Center for Educational Outcomes (NCEO) at the University of Minnesota as part of our standard process for years. Our content specialists, editors, and production staff are trained in these principles. We incorporate the NCEO principles of universally designed assessment in our item development process for paper-based and computer-based assessment materials. These principles include the following:

- Inclusive assessment population
- Precisely defined constructs
- Accessible, non-biased items
- Amenable to accommodations (e.g., braille and large print)
- Simple, clear, and intuitive instructions and procedures
- Maximum readability and comprehensibility
- Maximum legibility

Techniques that content specialists use and item writers have been trained to use in the development of items and stimuli include

- Use simple, clear, and easy-to-understand language in directions
- Keep the length of prompts and stimuli to the minimum required length, within the parameters of the item specifications
- Avoid sentences with multiple clauses unless needed due to defensibility of the item or part of permissioned stimulus
- Use a series of simpler, shorter sentences in place of longer, more complex sentences
- Use vocabulary and sentence structure that is below grade level for prompts and directions for mathematics and mathematical vocabulary that is at grade level
- Use vocabulary and sentence structure for prompts and directions that is at grade level when assessing reading skills
- Use common words
- Avoid the use of multi-meaning and ambiguous words, idioms, or jargon unless they are defined or part of the knowledge being measured

These materials are revisited periodically as new projects and materials are available for accessibility features and the content needed to support those features. Pearson will update training materials and internal checklists based on information in the final version of the PARCC Accessibility Guidelines, addressing the range of accessibility and accommodation features PARCC plans to have available.

Our Approach to Accessibility

Our approach to accessibility has been informed by a set of guidelines developed by Pearson in conjunction with the Center for Applied Special Technologies (CAST) for incorporating Universal Design principles specifically into computer-based testing. This document includes definitions and examples of how to apply the guidelines in a tiered system of test design and delivery. We have worked with companies such as the American Printing House for the Blind (APH), National Center for Accessible Media (NCAM), and Knowbility to refine practices for producing accessible content for students.

The use of technology in the assessments has provided more opportunities for accessibility beyond the standard practice of braille and large-print textbooks. The feature set which started with eliminate answer choices, flag items for review, highlight tool, notepad, and writing tools continues to expand. Features such as these do not have an impact on item development. Other features such as background/font color, text-to-speech, pop-up glossary, and magnification/enlargement require item developers awareness, and in most cases, content from item development to support those features.

Pearson has experience with those features. As we worked to implement color contrast on another project, it became evident that the some art blended in with the background color. This was solved by including a border around the art, typically white or black. Pearson worked with APH and NCAM in using PARCC's audio guidelines to create alt tags to support the creation of graphic description for use in text-to-speech. Some of the Texas assessments we produce include text-to-speech. Our art department has experience through GED, Virginia, and other projects of creating art that can be enlarged or zoomed, including the creation of scalable vector graphics. Some of the art used on the test form is permissioned. Raster graphics are typically the format available.

All mathematics items to be available online will have accessibility features applied to support text-to-speech (A13). We have used a combination of internal content staff and external providers to develop the text descriptions of graphics. All providers will receive training and complete a training set to verify a consistent application of PARCC's audio guidelines.

Pearson has a tool that provides a mechanism to incorporate alt tag descriptions and to produce APIP inclusion orders. Content providers can listen to the text-to-speech output and modify text in the inclusion orders to correct pronunciation. Upon contract award, Pearson would like to consult with PARCC on the expansion of PARCC's audio guidelines to include the application of alt tags and other descriptive information for technology-enhanced items. We envision this as an ongoing activity as new functionality types of items are included into development.

Other features that are to be applied to all items as appropriate include A9 (Magnification and Enlargement Device) and A11 (Pop-up Glossary). As part of the standard art production process, Pearson and its collaborative will produce high quality art as scalable vector graphics files, that can be enlarged onscreen as much as 400 percent. With the administration of the PARCC assessment on tablets, the screen size may limit what is a feasible enlargement for some larger graphics.

As we develop items, we will balance the size needed to clearly discern the information from the graphic with the size once enlarged. Universal Design techniques such as reducing clutter and increasing white space will be employed.

The Pop-up Glossary will be used primarily on the ELA/Literacy Assessment to define words within passages. PARCC requires the inclusion of authentic pre-published text that would require permission from the rights holder to make changes to the text. We will follow PARCC's established process where PARCC committees identify any words to be glossed. PARCC will approve all words that are glossed and their respective definitions.

In evaluating all passages and selecting words, content specialists will be cognizant of the vocabulary construct being assessed. Vocabulary is to be assessed in all passages, except narratives. Words with enough context for students to determine the meaning will likely not be candidates for glossing. It is anticipated that there will be few if any words glossed in mathematics or ELA/literacy items. When words are identified during the development process, the item will be edited to remove the words and replaced with more accessible language.

For the features in Table V.A.1.D.2.b, Pearson will create the content for 1 form per year for each grade/subject/assessment beginning in Spring 2015 to support the presentation accommodations noted in the table as needing item development. It is assumed that the same form will be used across the following options. PARCC and Pearson will work together to select forms that can be used for accessibility and accommodations.

- **P2. Braille.** Pearson will provide hard-copy braille tests for both ELA/Literacy and Mathematics using a provider approved by PARCC. We have worked with several different vendors over the years, as most state assessments provide this accommodation. The refreshable braille will use the same form as the paper, so the additional effort is related only to the delivery, not item development.
- **P3. Closed Captioning.** There is one multimedia passage on a PBA form, so we have assumed one passage per form in the costs for the production of the closed captioning. This includes a proofing process to verify synchronization of the video and the captions. The captioned text will comply with Word Wide Web Consortium (S3C) standards.
- **P4. Descriptive Video.** This narrated audio will augment the sound track of the video and provide descriptions of key visual element, such as describing a person walking on a dusty trail and their clothes are becoming dirty. This may not be evident from listening to the audio file, but may be a key piece to answer a question. Caution will be given to maintain the construct of the questions being asked. Permission to modify the soundtrack and an editable version of the soundtrack will need to be procured.

  Pearson will work with PARCC to determine how to implement this accommodation, if editable versions of soundtracks are not available or rights holders will not grant permission. No descriptive videos are included for mathematics.
- **P5. Tactile Graphics.** Paper-based raised surface graphics will be created to provide access to the visually impaired person using a computer-based accommodation such as text-to-speech or refreshable braille.

  In mathematics, there are graphics for which providing the level of detail needed to describe the graphic would change the construct being assessed. The tactile graphics can often be used to accompany the text-to-speech form.

In ELA/literacy, examinees using text-to-speech or refreshable braille may need to have the graphics provided as tactile graphics. Some art in ELA is not necessary for answering questions, so tactile graphics will only be provided for functional art. Pearson will adhere to the PARCC guidelines for tactile graphics. The parties will mutually agree upon the date by which the guidelines must be provided to Pearson.

- **P6. Text-to-Speech and Video of Interpreter for ELA/Literacy.** This accommodation has two deliverables. The first is a text-to-speech version of the ELA/Literacy Assessment. Pearson has experience following PARCC's audio guidelines for ELA/literacy. This will include alt tag descriptions of item and passage art. The second is an American Sign Language video of a human interpreter for the entire ELA/Literacy Assessment, including items, passages, directions, and response options.

  Pearson has worked with a vendor to provide sign language videos in the past and has performed an independent review of the video using staff training in American Sign Language. Pearson will adhere to the Sign Language Guidelines for developing the ASL videos. The parties will mutually agree upon date by which the guidelines must be provided to Pearson.
- **P7. American Sign Language Video for the Mathematics Assessment.** This accommodation for hearing impaired students will be managed in the same way as P6 above.
- **P8. American Sign Language Video of Test Directions.** This video will be of the initial test directions, only.

To fulfill the requirements listed above, Pearson will likely use a combination of internal staff and contracted staff. For example, Pearson has experience filming videos and producing audio files. However, we have limited internal resources certified as American Sign Language interpreters, so we would likely hire a consultant. Any consulting staff or organizations will be provided and trained to adhere to PARCC's Accessibility Guidelines and ASL Interpreter Guidelines.

Never a One-Size-fits-All Model

As stated in the PARCC Accessibility Guidelines, Universal Design focuses on the understanding that to increase access, assessment designers cannot use a one-size-fits-all model, but instead must open up opportunities for choice and create multiple alternatives for individuals. Accessibility needs to be addressed during item creation, not as an afterthought. All editing decisions should be made with a focus on creating items accessible to the widest population possible while maintaining the assessment of the construct.

This process starts with training content specialists and item writers, providing them criteria to evaluate their items with respect to Universal Design and accessibility and techniques to use to maximize accessibility of the item. All staff and reviewers of items, stimuli, art, and other assessment materials first need to understand the characteristics of the population and the content PARCC is intending to assess. The intent is to apply principles of Universal Design to reduce the need for accommodations.

Pearson will incorporate the following points, which are not currently part of our training, and provide them to item writers and content editors in the form of a checklist. During training, these key points will be discussed:

- **The item or task measures what it intends to measure.** The construct for PARCC is defined through the PARCC framework, evidence statements, and supporting item specifications. Items need to align to the construct of the evidence statement being measured. However, writers and editors should evaluate the item to determine if there is anything that would interfere with the item's ability to measure its intended purpose, such as the use of ambiguous words or linguistically complexity structure that is not needed.
- **The item or task respects the diversity of the assessment population and allows the full range of eligible students to respond to the item/stimulus.** The item should use accessible language with graphics, if they would assist in the understanding of the information in the item. The presentation or use of functionality should not interfere with what is being assessed.
- **Decisions will be made to verify that items and tasks measure what they are intended to measure for EL students with different levels of English proficiency and/or first language proficiency.** The items and tasks should use a sentence structure that is at grade level for reading and below grade level for mathematics. Multi-meaning words and passive voice often provide a challenge to EL students and should be avoided unless part of the construct being assessed.
- **All supports have been considered that may increase access while preserving the targeted construct or a range of students with disabilities.** PARCC has many functionality types of items that can be used to assess content, particularly in mathematics. The functionality type used should be intuitive to navigate for the question and may assist in making content more accessible. Some items may need to be adapted to provide support for students with certain disabilities, so an evaluation of whether the items allow for the use of braille, tactile graphics, or oral presentation. Notations in the item metadata should be made to identify any accommodations that would make the item accessible to a wider population.
- **Multiple means of presentation, expression, and engagement have been considered with regard to individual items/tasks for both SWD and Els.** There are often multiple ways to assess the construct. While PARCC desires a range of ways to assess content to allow for different learning styles, some ways are more accessible to SWD and ELs. Contrary to what one may think, items with some context are often more accessible that pure math items. The context can provide clues to how to solve the item that a purely symbolic one does not. Using contexts in mathematics that are familiar to students will reduce the construct irrelevant ideas student need to navigate through to answer the question and may add bias.

- **The item or task material uses a clear and accessible text format.** The principles of Universal Design include the use of simple, clear, and intuitive instructions and procedures that maximizes readability and comprehension.
- **The item or task material uses clear and accessible visual elements (when essential to the item).** The size and location of text, images, graphs, tables, or visual content should maximize legibility and not be distracting. Color can be used to differentiate information or provide emphasis online. Information within a graph should be clearly discernable, e.g., the placement of a geometric figure on or close to an axis may not be easily seen or result in the movement of axes labels. Scrolling should be minimized, particularly graphics needed to fit on a screen.
- **The item or task material uses text appropriate for the intended grade level.** The stimulus materials need to appropriate for the intended grade and meet the specifications of the PARCC assessment. Readability and text complexity rubric is used to judge the pre-published materials.
- **Changes to the format of an item will be considered to preserve the item/task meaning or difficulty.** The goal is to have items accessible to the largest proportion of the population. So as a final check, consideration is given whether a different format would be more accessible without changing the construct or the cognitive complexity of the item.

All item developers and item writers will be provided an electronic version of PARCC Accessibility Guidelines and PARCC's Linguistic Complexity Rubric. Items and passages are reviewed internally by accessibility and fairness experts trained in the principles of Universal Design and who become well versed in PARCC's Accessibility Guidelines. Checklists will be reviewed and updated with PARCC-specific criteria. Suggestions for edits will be provided to content specialists to evaluate with respect to the construct being assessed and experience with PARCC's application of the accessibility guidelines. Feedback will be provided to item writers along with follow-up training when necessary.

Requirement

V.A.1.E. Item Development Approach

1. Evidence-Centered Design
2. English Language Arts/Literacy Technical Approach
   - A. Materials for Elementary (Choose one grade from grades 3–5)
   - B. Materials for Middle School (Choose one grade form grades 6–8)
   - C. Materials for High School (Choose one grade from grades 9–11)
   - D. Optional Submission
3. Mathematics Technical Support

▪

Response Requirements for Section V.A.1.E.

a) Offeror's proposal shall include a response to the requirements specified in Section V.A.1.E.

b) Offerors shall provide ELA/literacy sample items

c) Offerors shall provide mathematics sample items

Response

Item Development Approach

This section will describe the approach to be used for the PARCC item development by Pearson and its subcontractors. As experienced PARCC developers, Pearson, ETS, and WestEd will bring the expertise in implementation of PARCC as well as knowledge from team members who have worked on other CCSS-based projects. The topics to be covered include the approach for Evidence-Centered Design (ECD) as well as other specifics for both ELA/L and Math. Samples for non-technology-enhanced items are also included in this section.

Fulfilling PARCC's ECD Framework

PARCC has expressed a strong commitment to using the principles, processes, and tools of ECD to build its comprehensive assessment system. As PARCC acknowledges, using ECD is likely to benefit students, educators, parents, other PARCC stakeholders, and the general public:

1. ECD provides a means for making the validity argument explicit and for documenting assessment design decisions and the rationales that support those decisions. Such evidence offers a robust basis for supporting the validity of score interpretations and test uses.
2. ECD design tools, such as Task Generation Models, enable assessment developers and subject matter experts to make transparent the expertise they bring to bear on the assessment design process so that this expertise can be disseminated for more widespread application.

The hallmark of ECD is the extent to which all the artifacts of the design process are linked in a hierarchical fashion. Beginning with the initial claims, moving to design tools (such as Task Generation Models), and ultimately to items and tasks themselves, each artifact is informed by the one that preceded it. However, as previously discussed, ECD is not a strictly linear process. In fact, ECD is meant to be iterative or cyclical in nature (Mislevy & Haertel, 2006).

PARCC has used the ECD process, as evident in the item development already done and in the language used in the RFP and the plan to provide ELA/literacy Task Generation Models and mathematics Design Patterns to support assessment development.

Using the Common Core State Standards prior to the development of any items, PARCC completed domain analysis, with hierarchical sets of claims the assessment system is intended to support being the primary output of that effort. PARCC had also initiated domain modeling by generating test blueprints, Model Content Frameworks in ELA/literacy and mathematics, and evidence statements that identify focal KSAs that will function as assessment targets. ELA/literacy evidence statements are leveled, in the sense that, for each grade, they reflect varying degrees to which corresponding standards are met.

For example, sample evidence statements provided in the PARCC RFP depict both partial and full satisfaction of specific ELA/literacy standards. The mathematics evidence statements are leveled by grade band, reflecting coherent progressions inherent in the Common Core State Standards.

Finally, PARCC used the process of articulating the Conceptual Assessment Framework by providing initial Task Generation Models in ELA/literacy and Design Patterns in mathematics. It is PARCC's express wish for these components of assessment design to be linked and aligned from start to finish to generate explicit documentation of evidence, transparency of assessment design processes, and comparability of test items and forms.

The contractors awarded this work will assume responsibility for continuing and extending the ECD process that PARCC has begun, including:

- Refining the Conceptual Assessment Framework by refining Task Generation Models in ELA/literacy and Design Patterns in mathematics
- Refining Item Generation Models in ELA/literacy
- Continuing assessment implementation by fulfilling item and task authoring sufficient to construct test forms according to PARCC blueprints

Proposed ECD Plan

Our proposed ECD plan for PARCC is based on our ECD expertise, as well as extensive experience in the application of ECD principles to different assessment situations by Pearson and its subcontractors. The RFP is clear that to meet PARCC's current vision this plan requires the following:

1. A thorough understanding of ECD and the use of ECD concepts and methods throughout the item and task development, review, and delivery process
2. A procedure to refine each of the ECD elements throughout the design process (e.g., claims, evidence statements, Task Generation Models, Design Patterns)
3. The incorporation of cognitive complexity into the ECD process and tools to support development of items and tasks that measure the full range of performance

To better meet these requirements, we propose the addition of two design considerations that we believe will add value and support PARCC in achieving its desired goals. They include the following:

1. The continued incorporation of additional attribute information into the current PARCC design tools (Task Generation Models, Design Patterns)
2. The continued integration of Universal Design into the ECD process and design tools

ELA/Literacy Technical Approach and Work Flow

While it may seem counterintuitive, much of our technical approach in ELA stems from our philosophical approach to content development. As former educators, the Pearson ELA item development team members are able to make careful considerations that are appropriate for each grade band. A teacher in the classroom rarely teaches topics in isolation, so it is realistic to consider that items for development might connect ideas across content areas while focusing on reading standards.

Within the 3–5 grade band, the team knows that the goal for elementary materials is that they parallel what happens in the classroom. Item development is based upon the questions, activities, and discussions that typically occur around reading materials, whether they be informational or literary. For example, the approach is not simply asking about what happened, but rather more of a link between ideas within the passage and other areas of learning. We will ask students to consider "how" and "why" rather than "who" to encourage deeper thinking and analysis. As required by most of the CCSS, item development requires an integration of ideas and pieces of the text rather than simple recall or restatement of what the author included.

In the elementary classroom, basic comprehension usually means very literal connections. More advanced and experienced readers are able to take the ideas of the text and begin inferring connections to other materials they have read or topics to which they are exposed in their reading or classroom experiences. Developing materials for the elementary classroom should begin with rigorous texts to verify that there is material within the passage to ask the more difficult and advanced questions for the experienced readers.

All too often, the focus is on comfortable reading for elementary passages, especially on assessments, because in the classroom it is necessary to provide a great deal of guided practice for struggling readers. Offering texts that permit all students to have a level of understanding will allow for challenge at all levels of reading expertise, even though the challenge is greater for struggling readers, just as it would be in the classroom. Some students will only grasp the basic plot and characters, while others will connect character motivation and subtle plot developments, and a few may understand how the text is organized or the theme or author's purpose.

In general, the PARCC item development philosophy has been to ask the questions that are begging to be asked' in each text. This unfolding of the text allows for varying levels of item development and rigor, embedding the CCSS among item types that use ideas and topics from the text into evidence-based items and technology-enhanced activities that truly parallel what classroom teachers are doing as they integrate the rigor and expectations embedded in the CCSS.

In middle school, as students acquire advanced reading skills and greater confidence in comprehending what they read, the need for challenging and appropriate assessment materials becomes even more important. At middle school the rigor in assessment is derived not only through texts of increasing complexity, length, and topic, but also through multifaceted standards that expect students to demonstrate a high degree of text control. Item development continues to be based upon the questions, activities, and discussions that typically occur around reading materials, but the focus moves to bringing to light subtle connections that are likely not readily apparent or obvious to most readers.

For example, the items challenge readers to analyze what they read or formulate responses to the text in a way that shows how information is conveyed, integrated, or elaborated in the text. We ask “how” and “why” rather than “who” to encourage deeper thinking and analysis. As required by most of the CCSS, item development requires an integration of ideas and pieces of the text rather than simple recall or restatement of what the author included.

While the middle school language arts classroom might focus on skills, genres, authors, and other content-specific topics in isolation in some activities, it is important to remember the expectation of CCSS to incorporate texts and topics from all content areas. Building reading expertise through texts that traditionally might have been used only in science or social studies classrooms, for example, is one attempt by PARCC assessments to broaden the range of materials students encounter in the assessment experience.

Particular focus on an author’s approach to a topic or how similar information might be presented in different formats or through different perspectives allows students to create links between what they are exposed to on the assessment and what they have studied in the classroom. Evaluating arguments and tracing claims in a text are typical skills of the middle school classroom and are incorporated into the assessment experience.

Similarly, examining how literary texts are structured or how specific facets of the text such as dialogue or humor are incorporated into a text allow students to discover greater meaning and relevance to what they read through analysis and summarization. Whether connecting ideas that are found explicitly or inferentially in the text, students are challenged to dig deep beyond the surface meaning of what they read and create logical and evidence-supported conclusions in the various types of texts and media in PARCC assessments at the middle school level.

For both middle school and high school, the use of strong literature and nonfiction passages is essential. Attention must be given to creating pairs of passages that fit the PARCC Task Models for literary analysis and triplets that fit the Task Models for Research-Based Tasks. Before passages are selected, our team must determine whether the potential passages can support both the required number of items and the required standards of the task model for which they are selected.

Nonfiction passages may come from non-traditional sources such as user guides and instruction manuals. Training of passage searchers is essential to obtaining very specific types of materials that are appropriate for each grade band.

At no level is there is a greater need for authentic texts than at the high school level. High school students need to show that they are truly college and career ready. This can best be accomplished by developing items aligned to passages that may mirror what a student would encounter in an academic or a professional setting. Students need the opportunity to demonstrate the ability to digest sophisticated texts independently.

In this vein, high school material will rely even more heavily upon non-fiction texts, and will also require students to conduct more analysis across genres. Non-fiction texts may draw more frequently from primary sources or seminal historical documents, as it is believed such texts are a more accurate representation of real-world text.

Items developed for high school will require students to engage in meaningful analysis of the text. Students must engage in higher order thinking to demonstrate comprehension of text (or texts). While students might be able to address theme at a lower grade level, at the high school level, students must be able to defend arguments or construct arguments around theme, the author’s approach to it, and the representation of theme throughout or across texts (or other relevant skills and topics).

When it comes to the technical approach Pearson takes, development begins with the selection of high quality authentic passages. Our team will select passages that comply with the PARCC Passage Selection Guidelines and provide sufficient context to allow students to provide evidence as captured in the PARCC evidence statements.

In addition, all passages selected for a particular Task Generation Model will provide context for the Task Generation Model focus skills. All selections will be translated into PARCC word templates, annotated with the readabilities, qualitative characteristics, and recommendations for grade level and cognitive complexity before being delivered to PARCC for a virtual Core Leadership Team review.

Each passage will be reviewed by internal bias and sensitivity experts as well as receive an extensive editorial review. Upon approval by the Core Leadership Team, passages will be prepared for a concurrent State Educator and external bias and sensitivity review.

Once passages have been approved for development, our team will begin item development. Each passage set will be developed with the appropriate number of evidence-based selected-response items, technology-enhanced constructed-response items, and prose constructed-response items as needed for the suggested use of the passage set (e.g.; EOY, or PBA).

Items will receive internal review for alignment to PARCC evidence tables, Task Generation Model, item selection guidelines, and accessibility and fairness reviews. In addition, Query and Test Interoperability (QTI) experts will review all technology-enhanced constructed-response items for compliance and viability within the QTI schema. Once items have passed the extensive internal reviews, they will be provided for a virtual Core Leadership Team review. Once items are approved by the Core Leadership Team review, they will be prepared for concurrent State Educator and external bias and sensitivity reviews.

The following figure shows the number of texts and items to be developed for field testing by Pearson and its subcontractors in year one of this contractand the number of texts and items to be available for operational use. PARCC recommended overages are reflected in the table. Development targets will shift, as needed, depending on the task counts in the bank and to support the mutually agreed upon item development plan.

			Task Types							
	Texts		Evidence-Based Selected-Response Items		Technology-Enhanced Constructed-Response Items		Prose Constructed-Response Items		Total Tasks by Grade and Component	
	FT	OP	FT	OP	FT	OP	FT	OP	FT	OP
Grade 3 PBA	14	10	62	24	23	10	17	6	102	40
Grade 3 EOY	15	10	77	36	26	16			103	52
Grade 4 PBA	15	12	70	30	23	10	17	6	110	46
Grade 4 EOY	15	10	77	36	26	16			103	52
Grade 5 PBA	18	12	70	30	23	10	17	6	110	46
Grade 5 EOY	15	10	77	36	26	16			103	52
Grade 6 PBA	18	12	70	30	23	10	17	6	110	46
Grade 6 EOY	18	12	77	38	26	14			103	52
Grade 7 PBA	18	12	70	30	23	10	17	6	110	46
Grade 7 EOY	18	12	77	38	26	14			104	52
Grade 8 PBA	18	12	70	30	23	10	17	6	110	46
Grade 8 EOY	18	12	77	38	27	14			104	52
Grade 9 PBA	24	18	100	45	24	15	24	9	156	69

Grade 9 EOY	24	18	111	57	39	21			150	78
Grade 10 PBA	24	18	100	45	24	15	24	9	156	69
Grade 10 EOY	24	18	111	57	39	21			150	78
Grade 11 PBA	18	12	70	30	23	10	17	6	110	46
Grade 11 EOY	18	12	77	38	27	14			104	52
Total Texts	**332**	**232**								
Total by Type			**1443**	**668**	**488**	**246**	**167**	**60**	**2098**	**974**

The following figure shows the number of texts and items to be developed for field testing by Pearson and its subcontractors in years two through four of this contract and the number of texts and items to be available for operational use. We anticipate that the number of technology enhanced constructed response items may increase; however, the test blueprints and task models for ELA/Literacy are very prescriptive. These counts are based on the current blueprint requirements. Pearson and PARCC may mutually agree to modify the blueprint requirements, and resulting item development targets.

			Task Types							
	Texts		**Evidence-Based Selected-Response Items**		**Technology-Enhanced Constructed-Response Items**		**Prose Constructed-Response Items**		**Total Tasks by Grade and Component**	
	FT	**OP**	**FT**	**OP**	**FT**	**OP**	**FT**	**OP**	**FT**	**OP**
Grade 3 PBA	30	20	132	52	48	22	36	12	168	86
Grade 3 EOY	30	20	165	78	55	35			220	113
Grade 4 PBA	36	24	150	65	48	22	36	12	186	99
Grade 4 EOY	30	20	165	78	55	35			220	113
Grade 5 PBA	36	24	150	65	48	22	36	12	186	99
Grade 5 EOY	30	20	165	78	55	35			220	113

Grade 6 PBA	36	24	150	65	48	22	36	12	186	99
Grade 6 EOY	36	24	165	82	55	30			220	112
Grade 7 PBA	36	24	150	65	48	22	36	12	186	99
Grade 7 EOY	36	24	164	82	58	30			222	112
Grade 8 PBA	36	24	150	65	48	22	36	12	186	99
Grade 8 EOY	36	24	164	82	58	30			222	112
Grade 9 PBA	42	30	175	79	56	27	42	15	217	121
Grade 9 EOY	42	30	195	100	69	37			284	137
Grade 10 PBA	42	30	175	79	56	27	42	15	217	121
Grade 10 EOY	42	30	195	100	69	37			284	137
Grade 11 PBA	36	24	150	65	48	22	36	12	186	99
Grade 11 EOY	36	24	164	82	58	30			222	112
Total Texts	**648**	**440**								
Total by Type			**2924**	**1362**	**980**	**507**	**336**	**114**	**3792**	**1983**

Math Technical Approach and Work Flow

Our collaborative team recognizes, understands, and supports the shift in assessment practices. Furthermore, we recognize and support the significant value−to students, teachers, parents, states, and the country− of transforming to the more sophisticated and innovative type of assessment system envisioned by PARCC.

PARCC has identified eight characteristics of innovative tasks, not all of which are necessarily presented in each individual task:

- Quality assessment of individual content standards
- Practice-forward
- Tasks assessing conceptual understanding
- Integrative tasks

- Fluency assessment
- Expressing mathematical reasoning
- Modeling/application
- Technology-enhanced tasks

The following is a discussion of each of the innovative task types that our content specialists will design. Overall, tasks should be aligned to the standard in question, be progression sensitive, minimize or avoid common drawbacks of selected-response, and embody important shift of CCSSM.

Quality Assessment Tasks Aligned to Common Core Standards

These tasks would include Type I with the exception of multiple choice or multiple select. This would allow for a more close alignment to the standard or evidence statement in question.

Practice-Forward

In addition to the content standards, the Frameworks include eight Standards for Mathematical Practice which will be assessed using the mathematical content. That is, tasks for which it would be unlikely or impossible to earn full credit on the task without engaging in the practice.

We will create tasks that will require students to generate a response and demonstrate at least one of the Standards for Mathematical Practice in conjunction with, but not exceeding the requirements of, the particular content standard being assessed. We recognize that these tasks should authentically connect assessment to classroom learning. Supplementing selected-response items with technology-enhanced and constructed-response items will help achieve this goal.

Tasks Assessing Conceptual Understanding

These tasks would include technology-enhanced items without the chance of guessing. Therefore, multiple choice and multiple select item types would be avoided as much as possible. These tasks will assess conceptual understanding, where the standards explicitly call for it. It might be that many or most of these are short tasks that are computationally non-intensive and easy to answer quickly, if the student understands the concept in the question, but difficult to answer at all if the student does not understand the concept.

Integrative Tasks

These tasks would include Type I or Type II tasks that involve the integration of several Common Core State Standards and as such are best coded to a cluster heading, domain heading or grade/course title, rather than a specific standard. This would allow the task to assess the compilation of mathematics understanding across various concepts.

Fluency Assessment

These tasks would include Type I at the grades 3–6 level to assess accuracy and reasonably fast computation. Fill-In-The-Blank and Equation Editor are often used to confirm that the fluency aspect is captured. Also, context and extraneous information is removed from these tasks to address the concept of straight computation.

Express Mathematical Reasoning

These tasks would include Type II and involve the students using reasoning. One important aspect of mathematical reasoning is to promote student's thinking rather than having them providing a computational method in which they used to solve a problem.

Modeling/Application

These tasks would include Type III and involve the students using modeling through the usage of equations or explanation on how to use models to solve the task. Pearson is diligently working towards more ways for students to interact with the modeling concept for the Type III tasks. The targets of these tasks will be to address one or more of the following steps in the basic modeling cycle:

- Identifying variables in the situation and selecting those that represent essential features
- Formulating a model by creating and selecting geometric, graphical, tabular, algebraic, or statistical representations that describe relationships between the variables
- Analyzing and performing operations on these relationships to draw conclusions
- Interpreting the results of the mathematics in terms of the original situation
- Validating the conclusions by comparing them with the situation, and then either improving the model or, if it is acceptable
- Reporting on the conclusions and the reasoning behind them

Technology-Enhanced Tasks

These tasks involve all the technology-enhanced items. Students provide a response by interacting with objects in the prompt. These prompts would:

- Approach assessments differently than the traditional multiple choice prompts
- Use their interactive formats to assess relevant knowledge, skills, and abilities indicated in the Common Core State Standards
- Be innovative and take advantage of the technology-enhanced formats. The functionality should be an integral part of measuring students' knowledge
- Require students to create responses rather than select responses

- Not include repetitive tasks where fewer tasks would sufficiently demonstrate mastery of the standard

PARCC in collaboration with Pearson and ETS have built spreadsheets at the evidence statement level that incorporates the probability statements from the test blueprints and attrition rates at committee review and data review. The basis of our entire item development will be driven by the use of these item development target spreadsheets provided by PARCC. Before beginning item development, Pearson will use these target spreadsheets to develop an internal item development plan to correlate with the expectations of the test design. These will be reviewed and approved by PARCC as discussed in V.A.1.A. We acknowledge that each assessment has multiple parts and each part specifies the types of tasks and standards eligible for assessment.

For example, Type I tasks appear on either the Performance Based Assessment, End-of-the-Year Assessment, or both, whereas, Type II and Type III tasks should only appear on the Performance Based Assessment.

Effectively developing innovative tasks and assessments necessitates using superior components. To deliver this for PARCC, the Pearson content development team will use a sound, established development process that includes careful recruitment and training of writers, extensive review and refinement of items, and close collaboration between PARCC and our staff.

The item development approach we plan to use to develop these CCSSM-aligned innovative tasks employs the following aspects:

- A collaborative development team well-versed in CCSSM and trained to develop CCSSM-aligned tasks using of the PARCC design and development inputs
- Content acquisition using a variety of approaches
- Internal content reviews
- Internal editorial, accessibility, Universal Design, and bias and sensitivity reviews

These four aspects of the item development process are further described below.

Our Collaborative Development Approach

Our content leadership team is well versed in the CCSSM and the PARCC specifications. We have developed our plans with the assumption that our collaborative team will fulfill the requirements of the RFP. To that end, we will use our network of content and assessment experts to provide this development at the scale required, coordinating our efforts for the extended development team to be well versed in the documents and tools at our disposal:

- CCSSM and its supporting materials
- PARCC Model Content Frameworks

- PARCC Task Models/Design Patterns
- Performance Level Descriptors
- Prototype Tasks
- PARCC's Accessibility Guidelines

Shortly after contract award, the content leadership team will review the documents and tools to determine if additional questions or clarifications are needed, as well as determining how to fit the operational development with Phases 1 and 2 of development. This will be important in building a foundation and common understanding of the operational assessment design and development targets for each assessment by task type.

The following figure shows the number of mathematics tasks to be developed for field testing by Pearson and its subcontractors in year one of this contract. The second table includes the annual number of tasks to be developed and available for field testing for years two through four. PARCC recommended overages of 20% for 1 and 2 point items and 33% for all other items are included. Development targets for specific Item Types may shift, as needed, depending on the task counts in the item bank and to support the mutually agreed upon item development plan. While the various Item Types are not "equal" in terms of the level of cost and labor required for development, Pearson and PARCC intend for any modification of development targets for specific Item Types to result in a relatively neutral cost impact, as mutually agreed by the parties.

	Type I			Type II		Type III		
	1-Point Tasks	2-Point Tasks	4-Point Tasks	3-Point Tasks	4-Point Tasks	3-Point Tasks	4-Point Tasks	Total Tasks by Grade and Component
Grade 3 PBA	20	5		6	6	6	3	46
Grade 3 EOY	85	13						98
Grade 4 PBA	20	5		6	6	6	3	46
Grade 4 EOY	70	20						90
Grade 5 PBA	15	8		6	6	6	3	44
Grade 5 EOY	60	20						80
Grade 6 PBA	20	5		6	6	6	3	46
Grade 6 EOY	65	18	3					86
Grade 7 PBA	20	5		6	6	6	3	46
Grade 7 EOY	60	20	3					83
Grade 8 PBA	25	3		6	6	6	3	49
Grade 8 EOY	65	13	6					84
Algebra I PBA	38			9	9	9	9	74
Algebra I EOY	79	42	14					135

	Type I			Type II		Type III		
	1-Point Tasks	2-Point Tasks	4-Point Tasks	3-Point Tasks	4-Point Tasks	3-Point Tasks	4-Point Tasks	Total Tasks by Grade and Component
Geometry PBA	38			9	9	9	9	74
Geometry EOY	72	45	14					131
Algebra II PBA	25			6	9	6	9	55
Algebra II EOY	48	30	9					87
Mathematics I PBA	25			6	6	6	6	49
Mathematics I EOY	48	30	9					87
Mathematics II PBA	25			6	6	6	6	49
Mathematics II EOY	48	30	9					87
Mathematics III PBA	25			6	9	6	9	55
Mathematics III EOY	48	35	6					89
Total Tasks by Type	**1044**	**347**	**73**	**78**	**84**	**78**	**66**	**1770**

	Type I			Type II		Type III		
	1-Point Tasks	2-Point Tasks	4-Point Tasks	3-Point Tasks	4-Point Tasks	3-Point Tasks	4-Point Tasks	Total Tasks by Grade and Component
Grade 3 PBA	40	10		12	12	12	6	
Grade 3 EOY	170	25						
Grade 4 PBA	40	10		12	12	12	6	
Grade 4 EOY	140	40						
Grade 5 PBA	30	15		12	12	12	6	
Grade 5 EOY	120	40						
Grade 6 PBA	40	10		12	12	12	6	
Grade 6 EOY	130	35	6					
Grade 7 PBA	40	10		12	12	12	6	
Grade 7 EOY	120	40	6					
Grade 8 PBA	50	5		12	12	12	6	
Grade 8 EOY	130	25	12					
Algebra I PBA	63			15	15	15	15	
Algebra I EOY	132	69	23					
Geometry PBA	63			15	15	15	15	
Geometry EOY	119	75	23					
Algebra II PBA	50			12	18	12	18	

	Type I			Type II		Type III		
	1-Point Tasks	2-Point Tasks	4-Point Tasks	3-Point Tasks	4-Point Tasks	3-Point Tasks	4-Point Tasks	Total Tasks by Grade and Component
Algebra II EOY	95	60	18					
Mathematics I PBA	25			6	6	6	6	49
Mathematics I EOY	48	30	9					87
Mathematics II PBA	25			6	6	6	6	49
Mathematics II EOY	48	30	9					87
Mathematics III PBA	25			6	9	6	9	55
Mathematics III EOY	48	35	6					89
Total Tasks by Type	**1791**	**564**	**112**	**132**	**141**	**132**	**105**	**2977**

As described above, we will use the ECD procedures and tools to define the required and variable attributes of the tasks that will be developed, and to measure the quality of developed tasks against those same task attributes. A discussion of the task model refinement process for mathematics is included in this section.

Content Acquisition Approaches

PARCC assessments require a large number of tasks to be developed. In addition to the task authoring capabilities within our respective organizations, we propose to acquire items and tasks by using independent organizations and contractors.

Pearson, ETS, and WestEd content teams will work with the organizations and contractors in this collaborative to have a direct effect on training quality, content creation, and content review. Along with the named groups, Pearson maintains relationships with potential consultants and independent contractors who can fulfill specific needs as they are identified.

Because of the work our organizations completed in Phases 1 and 2 of development, we already have a strong base of task writers trained on PARCC specifications and the CCSSM. We will continue to use our successful contractors while also adding new contractors trained on the PARCC requirements and the CCSSM. This will be a benefit to PARCC in that we already have a strong base and training materials developed.

Internal Review and Universal Design Review Process

The Pearson collaborative's process for content development includes conversations and problem solving between content specialists, editors, and accessibility specialists, as outlined in V.A.1.F. As noted previously in this section, the review process will be informed and guided by the PARCC design documents and tools. By completing these steps, our team can monitor the quality of development, identify issues, develop additional training when necessary, and prepare content for the formal reviews.

Sample Items

The following pages highlight Pearson's ELA/literacy and Mathematics samples.

ELA/L and Math Sample Items Redacted

Requirement

V.A.1.F. Contractor Review of Test Items/Tasks and Texts

Response Requirements for Section V.A.1.F.

a) Offeror's proposal shall include a response to the requirements specified in Section V.A.1.F.

Response

Contractor Review of Test Items/Tasks and Texts

As items and tasks are received by the respective organizations in the Pearson collaborative, the content specialists begin an internal review process that has been designed after decades of industry assessment experience and the experience that our content specialists bring from their own areas of expertise and classroom teaching experience.

The contractors have established processes to reflect best instructional practices to develop assets that meet the high standards that are expected by PARCC. Items and tasks will be scrutinized in a series of extensive reviews by various groups comprised of content, editorial, research, and accessibility/fairness review teams. Each team is responsible for their own particular area of expertise. This results in items/tasks that meet the strict requirements required of all PARCC assessment material.

Internal Content Review

The first step in this comprehensive process begins when at least two content specialists review all submitted assets, making any necessary revisions to enhance their quality from the perspectives of content, alignment to Common Core State Standards and evidence statements, and grade-level appropriateness. Assets submitted that contain significant issues, for example an item that does not match the evidence statement, will either be returned to the item writer for revision or rejected.

Some of the criteria that the content specialists will use some of the following criteria to verify the quality of the assessment material:

- Accuracy, meaningfulness, and level of engagement of item and task content/context
- Alignment with the Common Core State Standards and evidence statements
- Suitability/appropriateness of any context associated with the item
- Appropriate complexity and knowledge demands

- Accuracy and completeness of the associated metadata
- Suitability of any art requested and clarity of the art description
- Clarity of text and grade-appropriate language and any art being used
- Accuracy and completeness of the answer key/rubric provided with an item
- Any specific considerations related to the item's potential suitability for paper-pencil and/or computer based formats

The content specialists will collaborate to identify revisions and changes that will enhance the quality of the assets. For items with complex issues, the content specialists might consult with other experts to confirm that the best decision is made to resolve the issue and move the material forward in the process.

Fact Checking

If any assets contain information that is not self-evident or requires research because it is beyond the common knowledge held within that content area, these materials will be reviewed by a Pearson collaborative's research librarian. The research librarian will attempt to authenticate the accuracy of all information included in these materials with at least two independent sources. If this validation of information is not possible for some reason, the research librarian will notify the appropriate content staff so that an informed decision can be made.

Scoring Experts

Some item types such as constructed response may be scored by different scoring entities such as a performance hand-scoring center or an automated response engine. Where appropriate, these types of items will be reviewed by Pearson's scoring experts for their comments in how to develop an item to help make the scoring more valid and reliable.

Accessibility and Fairness Review

We will review all assets using the principles of Universal Design and PARCC's Accessibility Guidelines and Fairness Guidelines to provide the greatest participation and access for the widest possible range of students. In addition, we will develop items without creating impediments that may possible limit opportunities for success by members of various populations. The comprehensive process followed by the Pearson collaborative is documented in the Accessibility and Fairness.

Copy Edit

Pearson uses a series of checklists for reviewing the editorial content and structure of assets. Checklists are customized to reflect PARCC content stands and again for the style of items for each content area. In addition to the customized checklists, Pearson will work with PARCC to update the style guide to maintain consistency throughout asset development.

Our editors will review the following elements during initial asset development as well as after committee review meetings. In addition to these quality checks, editors will review assets for clarity, appropriateness of language for the grade level, adherence to style requirements, and conformity with acceptable asset writing practices.

Editorial Quality Checks During PARCC Asset Development	
Initial Editorial Review	**Element**
✓	Verify punctuation, syntax, and spelling
✓	Verify subject/verb and tense agreement
✓	Verify pronoun clarity, agreement, and consistency
✓	Verify the correct use of quotation marks
✓	Verify that wording used in sentences, captions, and direction lines is unambiguous and concise
✓	Verify that all options are parallel in length and grammar
✓	Verify correct passage quotations within items (as compared with the passage)
✓	Verify that item order reflects passage order
✓	Verify the correct content/spelling/grammar in graphics (e.g., tables, charts)
✓	Verify the correct content/spelling/grammar in art
✓	Verify that the asset meets PARCC style requirements (e.g., the correct formatting for emphasis words, distracters)

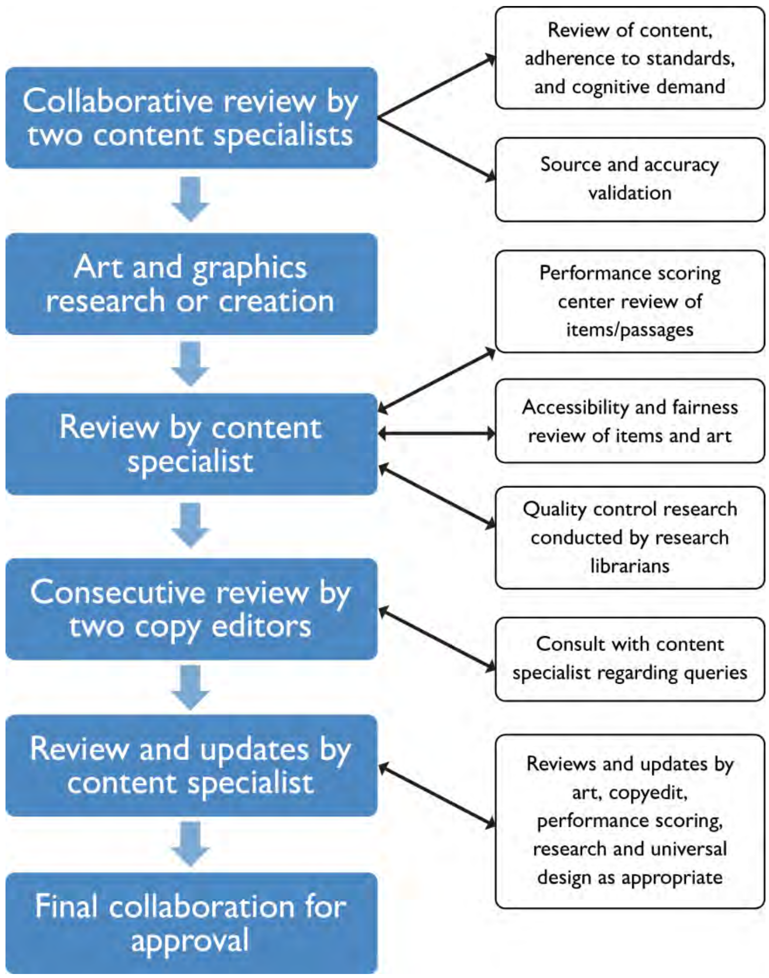
Two editors will conduct independent, consecutive reviews of the PARCC assets. The second editor will serve as a senior reviewer and will make any final decisions regarding editorial comments and queries. Editors maintain a library of books and online tools that they can use as further resources for checking the appropriateness of syntax and content. The editors will consult with Pearson content specialists to resolve any queries that arise related to the items/passages.

Final Content Review

After all steps have been completed as necessary, a content specialist will review each asset along with any associated feedback that was provided by the team to determine what revisions should be made to enhance the quality of the asset. After the asset is revised, the content specialist may choose to have an editor proofread the asset once again to perform a final quality check. Once all required revisions have been incorporated, the assets will now be ready for external Partnership review.

The following figure shows a flowchart of the internal item review process for most items.

Internal Review Process for Asset Development



Requirement

V.A.1.G. PARCC Process for Test Items/Tasks

1. PARCC Test Development Review Committees
2. Third-Party APIP Tag Technical Review
3. Test Development Review Meeting Requirements
   1) Partnership Review Meetings
   2) Test Development Meeting Requirements

Response Requirements for Section V.A.1.G.

a) Offeror's proposal shall include a response to the requirements specified in Section V.A.1.G.

b) The Offeror shall ensure that there is a description of the approach to meeting the following:

   i. Offerors will propose comprehensive solutions for conducting virtual test item and task reviews and in-person reviews.

Deliverables for Section V.A.1.G.

a) Deliver Text Review Meeting Plans
b) Deliver Text Review Meeting Agendas, Participant Lists, Meeting Notes
c) Deliver Text Review Training Materials
d) Deliver CLG Meeting Plans
e) Deliver CLG Review Meeting Agendas, Participant Lists, Meeting Notes
f) Deliver CLG Review Training Materials
g) Deliver SE Meeting Plans
h) Deliver SE Review Meeting Agendas, Participant Lists, Meeting Notes
i) Deliver SE Review Training Materials
j) Deliver B/S Meeting Plans
k) Deliver B/S Review Meeting Agendas, Participant Lists, Meeting Notes
l) Deliver B/S Review Training Materials
m) Deliver Editorial Review Meeting Plans
n) Deliver Editorial Review Meeting Agendas, Participant Lists, Meeting Notes
o) Deliver Editorial Review Training Materials
p) Deliver Third-Party APIP Report

Response

PARCC Process for Test Items/Tasks

This section is organized by deliverables required to best explain the approach to review meetings. Each of the requirements for this section per the RFP is covered within these various deliverable sections.

PARCC Test Development Review Committees

Review meetings involving key stakeholder groups are critical to the integrity of the item development process. The meetings will include a review of content development guidelines; ELA/literacy items, tasks, passages and their data; and mathematics items, tasks and their data. These meetings will include a combination of in-person and virtual meetings. Our program implementation team, content specialists, and psychometricians are experienced in coordinating and facilitating all meetings referred to in the RFP.

The following figure shows the sequence of reviews. Note that mathematics will start with the Internal Item/Task Review step.

Sequence of Reviews	
Stage	**Primary Task**
Internal Passage/Text Selection and Review	Passages will undergo many internal reviews by Pearson, ETS, and WestEd staff, including for content quality and appropriateness, copy edit, Universal Design and fairness, and fact verification.
PARCC Text Review	All passages that Pearson, ETS, and WestEd develop will be presented to the PARCC text review team for review and approval.

Sequence of Reviews	
Stage	**Primary Task**
PARCC Text Review Reconciliation and Validation	Approve edits and passages/text for item development.
Internal Item/ Task Creation and Review	Items and tasks will undergo many reviews by Pearson and our subcontractors, including for content quality and appropriateness, copy edit, Universal Design, accessibility and fairness, scorability, alignment to Task Models, and fact verification.
PARCC Core Leadership Review	All items and tasks we develop will be presented to the Core Leadership Team for review and approval.
PARCC Core Leadership Review Reconciliation and Validation	Approve recommended edits, rewrites, and tasks for State Educator review and/or bias and sensitivity review.
PARCC State Educator Review	Local education agency staff and higher education faculty will review 10 percent of the items and tasks for suitability of content.
PARCC State Educator Review Reconciliation and Validation	Approve recommended edits, rewrites, and tasks for bias and sensitivity review.
Bias and Sensitivity Review	Representatives of state educators and citizens from various backgrounds review items and tasks to confirm they are free from potential bias. Items/tasks approvals are eligible for field testing.
PARCC Bias and Sensitivity Review Reconciliation and Validation	Approve recommended edits, rewrites, and tasks for field testing.
PARCC Editorial Review	Local education agency staff will review 10 percent of passages, items and tasks to verify that there are no grammatical errors or text verbatim check errors.
PARCC Editorial Review Reconciliation and Validation	Approve recommended edits, rewrites, and tasks for field testing.
Internal APIP Tag Review	After the passages/text, items, and tasks have been through PARCC committee reviews, Pearson will apply APIP tagging and review the tagging for correctness.
Data Review	After field testing, educators, state content experts, psychometricians, higher education faculty, and accommodations/accessibility experts will review the item and task performance data.

PARCC Text/Passage Committee Reviews

Our content team is experienced in facilitating passage review committees. In the past year, we have facilitated over a dozen passage review committee meetings. We work with clients to develop processes to efficiently review materials according to the project's expectations.

Prior to each text/passage review meeting, both virtual and in person, we will document the meeting specifications. This documentation will specify the following meeting information:

- Meeting title
- If the meeting is in-person or virtual; for in-person meeting the location will be specified
- Conference call information
- Meeting participants
- Meeting facilitators
- Participant roles and responsibilities
- Meeting agendas, participant communications, and dining options for in-person meetings
- Requirements and list of training materials
- Requirements and list of reference materials
- Requirements and list of other materials and supplies

At the beginning of each committee meeting, our staff or the PARCC facilitators will provide thorough and effective training previously approved by Partnership Manager for committee members. PARCC training that is currently being used will be reviewed and revised as necessary. Text reviewers will be expected to provide feedback on issues such as the following:

- Is the text (print or multimedia) the appropriate length for the grade level?
- Is the topic of the text (print or multimedia) appropriate for the grade level?
- Is the complexity measure (print or multimedia) of the text appropriately coded?
- Does the text lend itself to the assessment of sufficient number of Common Core State Standards and evidence statements?
- Should any adjustments be made to excerpts? Is more of the original text needed in order for students to comprehend the text?
- Are there any words in the text that need to be glossed?
- Does the text respect the diversity of the PARCC assessment community?
- Will a segment of the PARCC assessment community be disadvantaged by the content of a text?

Many passages and some artwork will require prior permission to use on the assessment. The types of changes authors will typically agree to will also be included in the training.

We will work with PARCC to determine the criteria for the passage review and a checklist to help the passage reviewers.

During meetings, our staff will facilitate the review with the assistance of a PARCC designee, including logistical concerns, presentation of test content, recording of comments, and provision and management of required materials, including electronic equipment, reference documents, and other supplies.

Participants will be able to review the passages independently, through electronic display of passages, particularly multi-media passages, and then the grade group will discuss comments. Our content facilitator will reconcile with the PARCC designee upon completion of the review.

After a thorough editorial and content review of passages, we will provide any applicable changes to the passages for PARCC review and approval. Item development will begin after approval has been obtained.

PARCC Core Leadership and State Educator Review Committees

PARCC requires an efficient and effective process for test item and task review committees. We are experienced at developing and executing such processes and in facilitating effective and meaningful committee review meetings for both in-person and virtual settings.

Prior to each item and task review meeting, whether virtual or in person, we will document the meeting specifications according to PARCC requirements in the Meeting Plan Document as described in the committee review section above.

Prior to each meeting, Pearson will develop all materials, including training materials, to the Partnership Manager, at least one week prior to the meeting for approval.

At the beginning of each committee meeting, our staff will provide thorough and effective training for committee members. During the Core Leadership (CLG) review, these committees will review test items for adherence to the PARCC foundational documents, basic Universal Design principles, PARCC Accessibility Guidelines, selected metadata fields, and Style Guide.
Reviewers will provide feedback on issues such as the following questions:

- Does the item or task measure what it is intended to measure as laid out in the evidence statement, task model, and/or standards?
- Do the item or task materials have clear and concise directions indicating what the student is asked to do to answer the item or task?
- Does the item or task material provide sufficient information for the students to respond to the item or task?
- Are the items or task materials/stimuli visuals clear?
- Are the materials/stimuli essential to responding to the item or task?
- Are the items or tasks clear and concise?
- Is the content of an item or a task appropriate for the grade level?

- Is the content of the item or task correct?
- If the item describes a “real world” scenario, is that scenario plausible?
- Is the item metadata coded correctly for the specified fields?

The list of questions for the State Educator Review (SE) review differs slightly and includes one that asks if the item promote effective classroom instruction and assessment practices

During each meeting, our staff will coordinate and facilitate necessary aspects of the meetings, including logistical concerns, presentation of test content, recording of comments, and provision and management of required materials, including electronic equipment, reference documents, and other supplies. Our staff with work with the PARCC designee in each room to reconcile edits and with the designee at the CLG review for linguistic complexity review to capture the score and any concerns. For ELA/Literacy, the rooms will be configured by grade and by grade band for math.

PARCC Bias and Sensitivity Committee Reviews

Educators and community members will be asked to review items and task to confirm the absence of bias or sensitivity issues that would interfere with a student’s ability to accomplish his or her best performance. The objective is to provide items, tasks, and passages that do not unfairly advantage or disadvantage one student or group over another.

Committees of 16 grade band/content area members will meet periodically throughout the contract. In-person meetings will be scheduled for the first meetings of each development cycle and will move to virtual meetings for second review. All items and tasks will have been reviewed by the PARCC Core Leadership Team prior to being presented to the Bias and Sensitivity Committee.

At least one week prior to the review of the items and tasks, we will provide all training materials for all committee members to the Partnership Manager for review, and recommendations and approval. As with the other committees, Pearson expects to use the training materials that have been used for the other development effort and make minor adjustments as needed.

The training will include identifying characteristics that should be considered during the review, including PARCC’s Bias and Sensitivity Guidelines. Reviewers are asked to focus on the following two questions:

- Does the item or task respect the diversity of the PARCC assessment community?
- Will a segment of the PARCC assessment community be disadvantaged by the content of an item, or task?

Our staff will coordinate and facilitate necessary aspects of the meetings, including logistical concerns, test content presentation, comment recording, and provision and management of required materials, including electronic equipment, reference documents, and other supplies. Our staff with work with the PARCC designee in each room to reconcile edits.

During the bias and sensitivity review meetings, panelists will have the ability to review the items/tasks and stimuli in a similar format as they are presented to students.

PARCC Editorial Review Committees

PARCC requires all tasks to be free from editorial issues. PARCC editorial review committees will review 10 percent of items and tasks to verify that any editorial issues are corrected before inclusion in the PARCC item bank for future field testing. We are experienced at developing and executing such processes and in facilitating effective and meaningful committee review meetings for both in-person and virtual settings.

Prior to each editorial review meeting we will work with the Partnership Manager to select up to 10 percent of the items and tasks for this review. The PARCC editorial review committee participants will do their review in Pearson's item bank system. As with the other reviews, the committee members will view the items as the student would, and be able to vote and record their comments in the system.

As with other meetings, whether virtual or in person, we will document the meeting specifications according to PARCC requirements in the Meeting Plan Document.

Pearson will develop all materials, including training materials, to the Partnership Manager, at least one week prior to the meeting for approval.

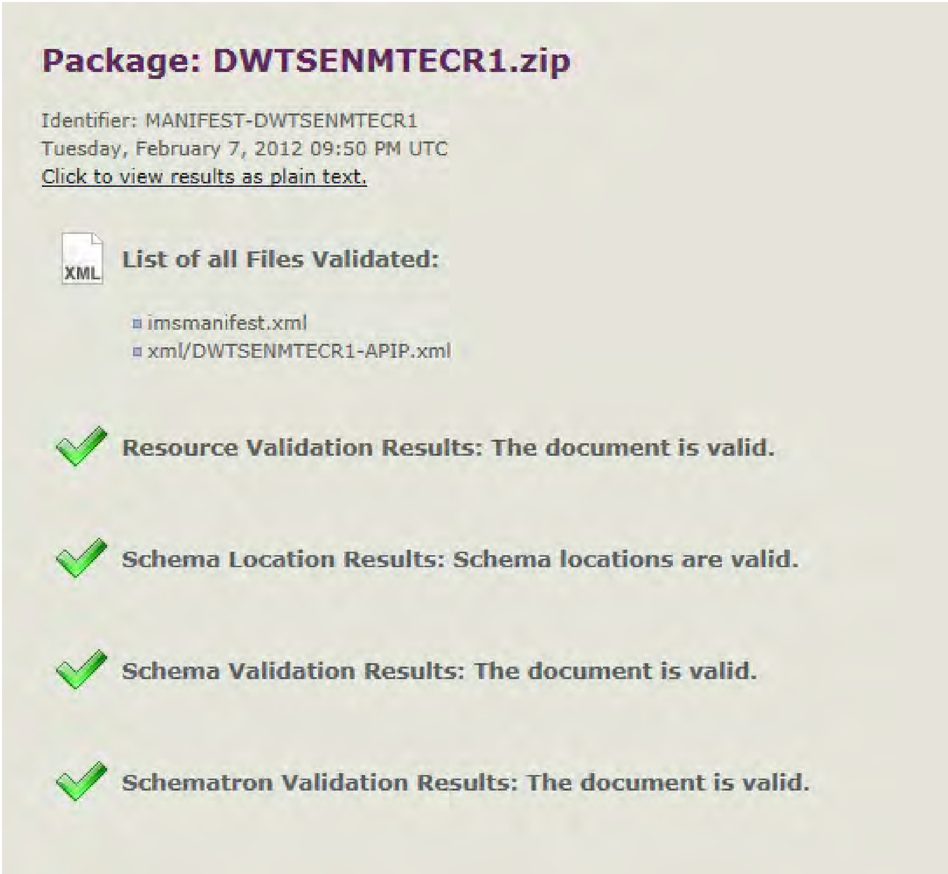
At the beginning of each committee meeting, our staff will provide thorough and effective training for committee members. Such training includes guidelines for reviewing the items and tasks for editorial issues.

During meetings, our staff will coordinate and facilitate necessary aspects of the meetings, including logistical concerns, presentation of test content, recording of comments, and provision and management of required materials, including electronic equipment, reference documents, and other supplies.

After the meeting, we will work with the Partnership Manager to determine if we would be responsible to conduct a complete editorial review of the remaining items and tasks not reviewed by the PARCC editorial review committee.

APIP Tag Technical Review by Pearson

After any final editorial issues have been corrected, Pearson will apply the APIP tagging and coding to the items. Pearson will complete validation using the APIP validator (http://validator.imsglobal.org/apip/). The following figure is an example of a report that verifies APIP compliance for an item.



Successful Verification/Validation. An example of how we validate APIP tagging and coding.

As PARCC has continued to innovate and expand the types of items being used, there have been a few cases where content required non-conforming extensions to the QTI XML. As one example, we have included the use of MathML with the inline choice (or drop down) interaction for several math items. This logical extension to QTI was required to support the use of equations or math symbols as choices in the drop down which is technically not supported by the QTI standard. All of these known extensions will be documented and excluded from any reporting as non-conforming items.

General Meeting Requirements

Security and Presentation

We share PARCC's expectations for security and presentation. Particularly with innovative items and technology-enabled accessibility/supports, a full evaluation of item content relies on the ability of item reviewers and stakeholders to experience, interact with, and respond to items as they will be presented to students. The reviewers will have access to computers and Internet during in-person reviews and we expect them to use their own computers during the virtual review. Reviewers will be able to access a secure, password-protected environment that has the items/tasks/passages and permits comments and votes to be entered.

All participants will be required to have a signed test security and non-disclosure form on file prior to being given access to secure materials. All materials used during in-person meetings will be secured, checked out when in use, and checked back in. We will work with PARCC to determine security procedures for both in-person and virtual meetings and include those procedures in training materials or correspondence with reviewers.

The items and tasks will be delivered for review in an environment comparable to the anticipated testing environment. Pearson's item review tool will provide reviewers with full item interactivity with what student's will see and experience on screen. Because the Pearson item previewer is specifically designed to handle the QTI/APIP encoded PARCC interactions, it is reasonable to assume that it would have a high degree of consistency with any test delivery system that is QTI/APIP-capable and support the PARCC item encoding.

Additional capabilities of the review tool include the ability to capture comments, view and manage metadata, review learning standards alignment, and inline editing. All of these capabilities are specifically designed to accommodate rapid revision and review of content. The tools will support all standard QTI interaction and PARCC specific items, from multiple-choice, to multi-prompt simulations, to graphing of math items, to mention a few.

Reference Materials

Pearson will have all reference materials approved by PARCC and identified in the Meeting Plan Document available to committee reviewers. The reference materials may include copies of the Common Core State Standards, Test Item Specifications, other PARCC foundational documents, basic Universal Design principles, PARCC Accessibility Guidelines, selected metadata fields, Bias and Sensitivity Guide, Style Guide, user guides to vendor's review sites, and other materials needed. For virtual reviews, Pearson will include in the Meeting Plan Document a list of materials and how any reference materials will be available to the committee reviewers.

Stipends and Substitute Reimbursements

Pearson will distribute stipends to all educators, both active and retired, who are not under contract during a review. Pearson will directly pay the qualified educators $150 a day for actual meeting days, not including any travel days.

Pearson will also distribute reimbursements to schools districts that require substitute coverage for active educators during a review. Pearson understands that if an individual receives a stipend, then a school district will not receive reimbursement for substitute coverage. Pearson will directly pay school districts $100 a day for actual meeting days and travel that does not occur on a weekend or state holiday.

Confidentiality Forms

Included in security and presentation plan above, Pearson will work with the Partnership Manager on a test security and non-disclosure form that meets PARCC requirements. Pearson will verify that we have a signed, active agreement on file before any reviewer views any text, passage, item, or task. Pearson acknowledges each agreement will exceed 12 months and committee members will need to complete an agreement annually.

Pearson will retain all signed test security and non-disclosure forms and provide copies to PARCC representative upon request.

In-Person Meetings

Pearson will work with the Partnership Manager to approve all locations before Pearson signs any contracts for in-person meetings. Pearson understands the location must be in PARCC state that is easily accessible by the participants and hotels must have the logistical requirements for the meeting, including Internet bandwidth and capabilities to handle the number of computers that will be used by the meeting participants.

During in-person meetings, each participant will be issued a computer (e.g. laptop, netbook, tablet) to access the online item review system and complete the review individually. We will also provide reference materials, as described above.

Following the individual review, items, and tasks that are not approved by all committee members will be projected and evaluated on a large screen, with particular attention given to those items or tasks that require large-group discussion. Our staff will record comments and suggested edits in the item review system, and also projected on the screen in the room. At the conclusion of the meeting, PARCC representatives will make final editing decisions and will reconcile meeting notes to inform post-meeting edits.

For all secure, hard copy material, Pearson will include a unique tracking number on each piece and require participants to sign in and out the material. The materials will be kept in a designated secure place when not used by the participants in the committee review rooms.

Pearson will include any conference call lines needed for in-person meetings in the Meeting Planning Document.

Virtual Meetings

Virtual review meetings will be conducted in two phases using the secure online item review system. The first phase will be an advance review completed individually by committee members. The second phase will be a group review convened via web conference.

In collaboration with PARCC, virtual meetings will be planned to address the potential for fatigue associated with extended virtual meetings and the need to accommodate reviewers across multiple time zones. The teleconference system used for virtual meetings will employ a secure virtual room system that will allow select reviewers to observe and interact with multiple grade/subject committees.

During the individual review, committee members will be given access to the items and tasks in the secure item review system for at least five business days during which they will review, vote, and comment on all assigned items. Prior to the group review, Pearson facilitators will run reports of the reviewers' votes and comments. Using these reports, Pearson will identify which text, items, or tasks the committee members did not reach complete agreement on and identify those during the group discussions.

During the group review, the meeting facilitator will show each item and associated comments, with particular attention given to those items or tasks that require large-group discussions. Comments and suggested edits will be recorded in the item review system and participants will be able to view what is being recorded via the net meeting.

To promote security for these virtual meetings, test security and non-disclosure agreements will be collected from all committee members that Pearson does not have an active one on file for prior to each meeting or being given access to the review site. These will be stored at Pearson for the length of the contract. Committee members will agree not to share information with anyone about items or about access to the items stored in the item review system. Additionally, reviewers will not duplicate or copy any material stored in the item review system. The item review system is hosted on a secure server with access that is password-protected.

Preparation for Item Review Meetings

Prior to all meetings, whether in-person or virtual, Pearson will work with the Partnership Manager to plan and document meeting requirements to verify that each participant has a successful and positive experience.

Meeting Plan Document

Pearson will provide the Partnership Manager with a Meeting Plan Document for each meeting, whether in-person or virtual, at least eight weeks before the start of the meeting. This document will include:

- Meeting title
- If the meeting is in-person or virtual; for in-person meeting the location will be specified
- Key dates (meeting dates, communication with participants, etc.)
- Conference call information
- Meeting participants
- Meeting facilitators
- Participant roles and responsibilities

- Meeting agendas, participant communications, and dining options for in-person meetings
- Requirements and list of training materials
- Requirements and list of reference materials
- Requirements and list of other materials and supplies

Logistics

Pearson welcomes the opportunity to provide the meeting services for the work described within this RFP. Each year Pearson is responsible for planning and coordinating hundreds of meetings, including item and data reviews, and other meetings required by this contract. For the PARCC Operational Assessment, we propose using our meeting planning group led by Suzie Nielsen, a certified meeting planner.

Pearson will provide our proposal for approval for conveniently located meeting spaces within PARCC states while taking into account the availability of restaurants located within or nearby the meeting facilities. For costing purposes, we have assumed that meetings will be located in centrally located or hub cities that are easy to get to for the majority of PARCC members.

In advance of in-person meetings, the Meeting Planning Document will include details such as the number of large gathering rooms, breakout meeting rooms, equipment, and sleeping rooms to name a few. Our meeting planning group has established strong working relationships with many large hotel chains.

Meeting Participant Confirmation and Communication

Pearson uses Cvent, an online meeting management website, to make it easier for meeting registration by automating critical parts of the invitation and registration processes. With Cvent, we can set up meeting places and times, provide information about location and content, market the events via email, and register participants— all online. Cvent also has reporting capabilities to support easy access for anyone needing to gain insight into status and tracking of meeting participants.

Cvent permits meeting managers to personalize communication to a target audience, set specific times for delivery of invitations and reminders, and track whether participants are opening and responding to email invitations and reminders. Registration and confirmation information will be available to PARCC, the Partnership Manger, and Pearson on-demand. Authorized users would be able to access the registration process anywhere, anytime. This hosted solution also eliminates additional website or technology support costs.

Prior to each meeting, Pearson will work with the Partnership Manager to identify a list of potential meeting participants and Pearson will handle contacting those participants for the meetings. Cvent will also be used to follow up with as needed after each meeting.

Meeting Materials and Meeting Training Materials

Pearson will develop/revise all meeting and training materials listed in the Meeting Planning Document. Pearson will include a timeline in the Meeting Planning Document to:

- Pearson develops the materials
- Pearson sends to the Partnership Manager to review
- Pearson makes recommended revisions from the Partnership Manager
- Return to the Partnership manager for approval

This timeline will put a plan in place to support review and approval of the materials at least one week prior to the start of the meeting.

Pre-Reviews

All items, tasks, and/or passages will be available at least to participants at least five business days prior to the start of a virtual review via Pearson's secure item review system. Participants will be provided with instructions, user ids, and passwords in the communications delivered through our Cvent system. The communication will include a toll free number to contact our Help Desk with any issues. The Help Desk assistance will be available a minimum of 9:00 Am to 7:00 PM EST Monday through Friday, excluding some holidays each day of independent reviews and group discussions.

Meeting Requirements

Pearson will provide staff in each review room that are both experienced in facilitating review meetings, as well as familiar with the text, items, and tasks being reviewed. In addition, Pearson will provide sufficient support staff for all logistics.

In addition, we will use Pearson's meeting planning group led by Suzie Nielsen. Suzie is supported by four full time employees. In 2011, the meeting planning team provided support for over 20 separate programs totaling nearly 300 meetings, over 800 meeting days, and over 14,800 participants. The meeting planning groups provide support and expertise to the program management team, which typically manages and provides direction for multiple aspects of an assessment program.

The program management team will work with the Partnership Manager and our meeting planning group to facilitate pre-event activities to verify that they are successfully completed prior to each event, and provide additional meeting coordination support during large meetings.

Post-Committee Process

Pearson facilitators will provide the Partnership leads with the changes and ideas recommended by the committee for their final approval. Pearson will include these final decisions in their notes.

Pearson will provide the Partnership Manager a list of the items and tasks that require edits and rewrites no later than three business days after reconciliation ends.

After each meeting, Pearson will also forward the Partnership Manager reports out of their item bank review system that captures feedback from the committee reviewers.

After all decisions are finalized and recorded, our content staff will incorporate requested edits and submit the revised items, tasks, and passages to our Copyedit group to verify that all edits have been applied as requested. Pearson's research, art, and Universal Design teams may also be included in the post-committee review process to assist in making revisions to specific items, tasks, and passages as needed.

All changes made during the editorial process will be captured in our electronic workflow system. The accuracy and appropriateness of edits will be verified by the content specialists, and we will deliver the revised items and tasks to PARCC no later than ten business days after reconciliation ends for validation. Pearson will also schedule a debriefing after the Partnership has completed their validation.

Test Development Review Meeting Requirements

Within 45 days after the contract is awarded, Pearson, ETS, and WestEd staff will meet with PARCC representatives to review test development and meeting requirements. This is one of the two Test Development planning meetings that will be held at Pearson each year.

Partnership Review Meetings

Below is a summary of the in-person content development meetings planned for year one (contract award–June 2015). The "Travel and Meetings" document attached to this Proposal includes similar information for years two through four of the contract for all virtual and face-to-face meetings required by the project. Some meetings, such as data review, do not occur in the first year of the contract and are not listed below but are part of our year two assumptions.

Meeting	Year 1 In-Person Meetings	Year 1 Virtual Meetings	Meeting Duration Per Meeting (Days)	Number of State Participants Per Meeting	Number of PARCC Staff Attending	Number of Contractor Staff Attending	In-Person Meeting Location
Text Review	2	0	5	76	6	17	Hub City
Core Leadership Review (ELA)	2	0	5	61	6	18	Hub City
Core Leadership Review (Math)	2	0	5	61	6	18	Hub City
State Educator (SE Review) (ELA)	2	0	4	76	6	18	Hub City
State Educator (SE Review) (Math)	2	0	4	76	6	18	Hub City
Bias/Sensitive (B/S) Review (ELA)	1	0	4	52	6	15	Hub City
Bias/Sensitive (B/S) Review (Math)	1	0	4	52	6	15	Hub City
Editorial Review (Math and ELA)	1	0	5	13	2	14	Hub City

Note: ELA/Literacy and Mathematics meetings are counted separately, but planned to happen concurrently.

General Meeting Requirements

Maintaining Item/Task/Text Security

We share PARCC's expectations for maintaining item/task/text security. It is Pearson's responsibility to make sure security is maintained prior to, during, and following the in-person and virtual content development meetings. While it is required that committee members have access to the secure materials to review test items, tasks, and texts, and make sure that they meet the PARCC requirements, we will perform the following step to maintain the security of PARCC's assessment content:

- All participants will be required to sign the PARCC test security and confidentiality (non-disclosure) form. Participants that attend multiple meetings will be required to sign a new form once a year. Prior to the content development meeting, the partnership manager will provide the form to Pearson, who will manage the distribution and collection of the forms and keep records to make sure that all reviewers have signed. We will maintain the original copy of the forms until the end of the contract. Copies of the signed forms can be made available to PARCC representatives upon request.

- During in-person and virtual reviews, participants will review content online, through a secure, password-protected environment that has items, tasks, and texts and permits comments and votes to be entered.
- Reviewers will be assigned a set of items, tasks, or passages to review, by grade/subject combinations. This will minimize the exposure any one reviewer has to the secure PARCC content.
- For in-person meetings, print copies of the test items, tasks, and reading passages being reviewed, if reviewed in hard copy, will be provided to each participant in a security-controlled notebook. The materials will be secured, checked-out when in use, and checked back in during use. Secure materials will not be permitted to leave the meeting room, cannot be photocopied, and will be stored in a locked location before and after meetings. From Pearson, we will ship materials to and from the meeting site using overnight shipping with tracking capabilities. Any materials that can be discarded after a meeting will be shredded.

Providing Necessary Reference Materials

For the in-person meetings, each participant will be issued a laptop to access the online item review system and complete the review individually. Reference materials as determined and documented in the Meeting Planning form will be provided in the agreed upon format.
All supporting materials such as audio/visual equipment, meeting supplies such as flip charts, note pads, pens, pencils, and other office supplies will also be provided for the in-person meetings.

Providing Stipends and Substitute Reimbursement

As required in the RFP, for any text review meetings, we will provide participants with a stipend of $150/meeting day unless they are under contract for teaching, in which case we will provide their school district with a substitute reimbursement of $100/meeting day. For costing purposes we assumed that 50 percent of the participants would receive stipends and district reimbursements would be provided for the other 50 percent of text review meeting participants.

For the state educator and bias and sensitivity review meetings we assumed that all of the participants attending in July/August would receive the stipend. Pearson will provide substitute reimbursement for participants attending these meetings in other months.

General Meeting Requirements for In-Person Meetings

In-person content development meetings will be held in a PARCC state hub city. For cost estimating purposes we assumed that hub city meetings would be held in Chicago, IL, or another PARCC hub city with similar GSA rates. For participants’ convenience, we plan to hold the meetings in a hotel near the airport so that shuttle service will be available for most attendees, reducing the need for rental cars and the associated cost. Because some meetings are a week long, we realize the need to select hotels that have onsite as well as nearby food options for dinner.

Given the size of the meetings—close to 100 total participants including contractor staff—we will select hotels that have demonstrated capability to handle this number of computers and with sufficient Internet connection. All meeting costs, including meeting rooms, audio/visual equipment, computers, and meals (continental breakfast and lunch) are included.

We will provide participants with a per diem to cover the expense of dinner, transportation to and from the airport (if needed), mileage, tolls, and other travel related expenses. Per diem rates will be based on the GSA per diem rates for the city in which the meeting is held. For costing purposes, our proposal assumes Chicago GSA per diem rates.

Given the large number of in-person meetings included under the proposed contract, Pearson will put together a comprehensive meeting plan at the beginning of each year for PARCC approval that will summarize the location, recruitment needs, and other key details for each meeting to be held during the year. Pearson will use this plan to recruit participants, and will provide instructions for booking travel, and pay reimbursements after the meetings.

General Meeting Requirements for Virtual Meetings

We expect that participants for virtual reviews will have access to computers and the Internet during virtual reviews. Reviewers will be able to access a secure, password-protected environment that has the items/tasks/texts, and permits comments and votes to be entered. A toll-free conference line will also be provided for any virtual meetings, including WebEx as needed.

Pre-Meeting Requirements

In addition to providing PARCC with an annual plan summarizing all in-person and virtual meetings that require recruitment and other types of logistical planning, we will provide the Partnership with a meeting planning document eight weeks before each meeting. The meeting planning document provides additional information that is specific to that meeting, including the following:

- Title of meeting
- Date and time
- Conference or WebEx information, if applicable
- Meeting agenda
- Participant names, including meeting facilitators
- Requirements for training materials
- Requirements for reference materials and supplies

For in-person meetings the meeting planning document will also specify:

- Meeting location (city, hotel)
- Instructions for booking travel
- Nearby dining options

Pearson has developed user-friendly tools to collect and disseminate event details. Our preliminary meeting specifications form helps us document and track your needs. The following figures show sample specifications forms Pearson will use to collect general information on each event.

PEARSON

Preliminary Meeting Specifications	
General Information:	
Program name	
Project number	
Cost center	
Name of meeting (how to post on site)	
Dates and times	
If flexible, list alternate dates and times	
Total number of attendees (include Pearson employees)	

Services Needed: (Mark all that apply):

Meeting arrangements	
Online registration, nominations, surveys (Cvent)	
Materials/supplies	
LCD projectors	

Forms:

Honorarium and amount	
Substitute teacher and amount	
Expense reimbursement (indicate mileage rate and limits on meals)	

Non-disclosure agreements	
Other	

PEARSON

Facility Information:

Note: Room and tax, catering, and audio/visual will be charged to Pearson. Guests will be responsible for incidentals unless otherwise specified.

Meeting location (city and state) Meeting site (if known or TBD) Hotel (if known or TBD)	

Room block by day and date				These counts are used in contract with the hotel, so please be as accurate and realistic as possible				
Day Date	Mon	Tues	Wed					
Number of sleeping rooms								

Do you need a general session room? (Indicate which days and setup)	
Do you need breakout rooms? (Indicate how many each day and setup)	
Number in each breakout room	
Do you need a secure office?	

Audio/Visual needs (Also list any other equipment needed)	

Food/Beverage (indicate Y/N) Breakfast AM break Lunch PM break Other	
Special dietary considerations (allergies, vegetarian, etc)	

Detailing Facility Needs. Our process gathers general meeting requirements in an efficient format so we can start planning for the PARCC Operational Assessment.

We will use this information to develop daily meeting requirements including:

- Set-up time
- Meeting times
- Meeting room name
- Room set-up requirements (classroom, u-shaped, etc.)
- Audio/visual requirements
- Food/beverage requirements

The following figure shows a sample of detailed meeting specifications Pearson will use to plan and deliver services for the PARCC Operational Assessment.

Meeting Planning Detailed Program Specifications

Wednesday, September 20- CONFERENCE DAY 1

WD-1	7:00 a.m.– 11:59 p.m.	6:30 a.m.	Conference Room 1	Staff office (DNP) Storage/ work room	Staff office (DNP)	Existing boardroom set Chairs around table Power strip for table 1 6-ft table against wall for storage with no chairs 1 large trash can (2) 6-ft tables in front room for registration with (2) chairs Note to hotel: We will need six keys made for this room. Secure room. DO NOT clean	N/A	INTERNET ACCESS codes provided	WD-1

						room.			
WD-2	7:30–9:30 a.m.	7:00 a.m.	Conference Room 2	Post-general session	95	Rounds of 8		N/A1 8’ x 8’ tri-pod screen LCD projector stand Extension cord Surge protector Podium w/microphone **Pearson to supply the following equipment:** LCD projector laptop	City Scramble (set as buffet) outside the room PM Break @ 2:00 p.m. cheesecake from lunch
WD-3	9:00 a.m.–6:00 p.m.	9:00 a.m.	Conference Room 3	Post-item writing English III (Group A)		U-shape for 24 6-ft table, skirted, at front of room 6-ft table, skirted, on side wall for storage Small round table at back of room with 2 chairs	Standard breakout room set 1 8’ x 8’ tri-pod screen LCD projector stand Extension cord Surge protector **Pearson to supply the following**		

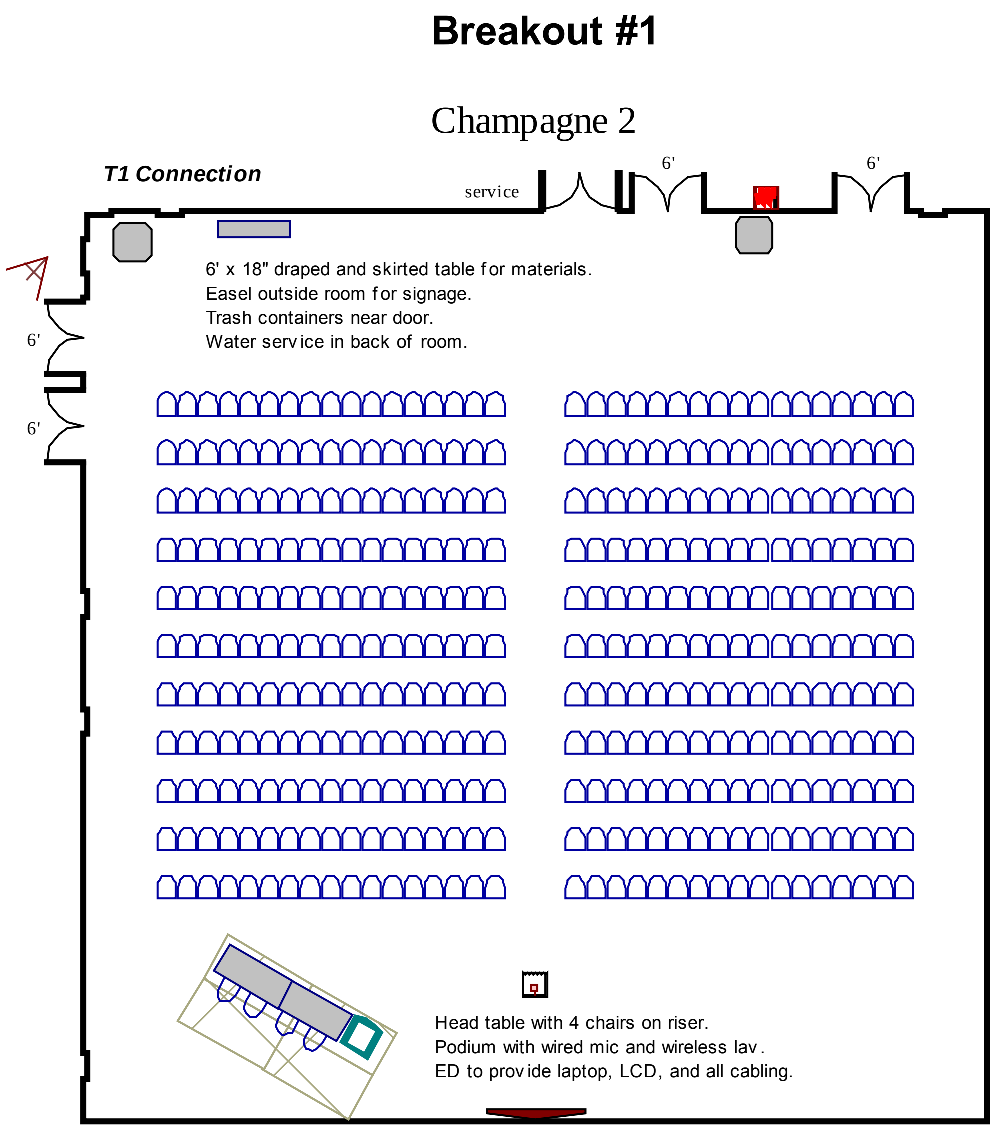
							equipment: LCD projector laptop		

Detailed Meeting Specifications. For PARCC, a key benefit of these specifications is that they provide daily meeting requirements including set up time, meeting room name, media requirements, food/beverage requirements, and more.

Meeting Room Setup Diagrams

Meeting facility personnel find that our meeting room setup diagrams provide useful visual references to detailed meeting specifications. To verify that our meeting specifications are clear and easy to follow, Pearson uses Meeting Matrix® software to download meeting room layouts from participating event venues. This allows us to configure detailed meeting room layouts that include the location of a screen, lectern, audio/visual cart, microphones, chairs, and tables as needed.

The following figure shows a sample meeting room floor plan.



Arranged to Fit PARCC Needs. Using Meeting Matrix, we can set up room layouts in a wide variety of configurations.

The following figures are sample forms that Pearson will use to collect information about meeting materials required for each meeting.

PEARSON

Materials Specifications	
General Information:	
Program Name	
Project No.	
Cost Center	
Program Team Contact	
Email	
Phone number	

Name of Meeting	
Dates and Times	

If more than one meeting, include order of meetings for printing and shipment purposes.

Materials:

Address you would like materials sent (folders, supplies, etc.)	
Name of person receiving materials (Pearson on site contact, hotel sales manager, etc.) Attn: Polly Planner	
Date you would like materials to arrive	

Note: All materials will be charged back to your project/ program.

Supplies:

Do you need supplies? (pens, rubber bands, pencils, scissors, etc) **Please list**	
How many supply boxes will you need?	
Do you need name badges?	

Do you need name tents?	
Do you need sign-in sheets?	

Please include a list of all names requiring name tents and/or badges.

PEARSON

Forms:

Indicate which forms will go in the folders and the number of copies needed.

Please include all forms in the table below

Honorarium and amount	
Expense reimbursement (indicate mileage rate and limits on meals)	
Substitute teacher and amount	
Non-Disclosure Agreements	
Participant List	
Other	

Folder:

How many folders will you need?	
Please specify colors that are acceptable.	
Would you like them labeled?	
If so, how would you like the label to read?	

Folder Layout:

Please specify how you would like the folders stuffed.

Folder Pocket	Paper Color	Black or Colored Ink	1 or 2 Sided
Left			
Right			

Determining Material Needs. Pearson meeting planners will use forms such as these to document the materials needed for PARCC committee meetings.

At the beginning of the year and/or prior to each meeting, PARCC will provide a list of potential state participants that includes contact information Pearson can use to invite educators to the review meetings. Invitations with logistical information will be sent at least six weeks prior the meeting. Pearson will be responsible for following up with potential participants to confirm attendance and send additional invitations as required.

At least one week prior to the meeting, Pearson will provide the partnership manager with materials to be used during review meetings. All items/tasks/texts will be available for online review for virtual meetings at that time. Although the PARCC item banking system has not been contracted yet (open Technology Bundle procurement), Pearson will provide a secure system as mutually agreed upon by Pearson and PARCC staff for the partnership manager to use for accessing and reviewing items prior to the meetings.

Meeting Requirements

Pearson, ETS, and WestEd will provide knowledgeable and experienced content and program management staff for the review meetings. As described in the Program Management Section of the proposal, Pearson will dedicate a full time event planner to the PARCC Operational program.

Ms. Deandrea White will attend all content development meetings to provide administrative support and will be responsible for all logistics, including questions about reimbursement and travel. She will work directly with the hotel prior to and during the meeting to make sure requirements are met, including room set-up, meal planning, Internet access, and other logistical considerations. She will be supported by program management staff that will be taking notes, and participating in separate subject/grade groups.

Content specialists from Pearson, ETS, and WestEd will facilitate the content development meetings. Content leads will be available if questions or concerns arise. The names of the contacts will be provided in the planning document.

A PARCC representative will chair the content development meetings, with additional PARCC representatives also attending each meeting.

Test Development Meeting Requirements

The items and tasks will be delivered for review in an environment comparable to the anticipated testing environment. Pearson's item review tool will provide reviewers with item interactivity. Once the system is operational in 2015, reviewers will be able to access online tools, embedded supports, and dynamic tailoring of item content and appearance to match what students need.

Additional capabilities of the review tool include the ability to function offline or in a secure networked environment, comment capture, global formatting controls, metadata display, a standards alignment viewer, inline editing, and dynamic reloading of digital item content to accommodate rapid revision and review of content including interactivity. The tool supports a wide variety of XML-based item types from multiple-choice to multi prompt simulations to graphing, and the review tool's extensible interactivity framework makes it easy to add in support for new XML-based item types.

Items will be grouped so that items/tasks that will appear together on an assessment, such as those associated with the same pairs or groups of texts, are located in the same review. Prior to developing items, proposed passages will be reviewed at a text review meeting.

Each test item, task, and reading text will be displayed on a screen and changes will be recorded for test items, tasks, and texts (if appropriate) onscreen throughout the review process.

Following each meeting, contractor staff and PARCC leads will participate in a reconciliation session to evaluate changes and ideas recommended by each committee. The Partnership will have final approval of all changes to be made. Pearson will compile the feedback and decisions and provide to PARCC.

Test Development Reconciliation Debrief

At the conclusions of each test development meeting, Pearson and its subcontractors will reconcile decisions from committee members and compile a list of decisions, including acceptance, minor edits, major rewrites, and rejections.

Staff will revise items in the item bank system as requested by committee members. Once that summary is delivered to the Partnership and edits have been applied within 10 days as requested by the RFP, individuals who have been appointed to reconcile will be able to go into the system to verify edits in a set five-day period.

Once completed, Pearson and its subcontractors will meet with Partnership staff virtually to debrief. A test development reconciliation process and debrief will occur by content area and grade span after each committee review meeting.

Requirement

V.A.1.H. Test Development Participants' Travel Costs

Response Requirements for Section V.A.1.H.

a. Offerors must include budgeted amounts in their Price Response/Reply and Budget Worksheets to cover the all reimbursable costs for all participants' air travel, car rental, lodging, and miscellaneous expenses such as parking, tolls, vicinity mileage, and other costs Offerors identify as necessary to participants' travel. Participants will also be reimbursed for one meal allowance and/or partial per diem.

b. Offerors will include budgeted amounts in their Price Response/Reply and Budget Worksheets to cover all reimbursable costs, including but not limited to costs of meeting rooms, and all supporting materials such as audio/ visual equipment and meeting supplies such as flip charts, reference materials, note pads, pens, pencils, and other office supplies, as well as costs related to shipping and/ or producing materials onsite, the disposition of secure materials at the conclusion of the meeting, and other costs Offerors identify as necessary to successfully conduct the meetings.

Response

Pearson will use Cvent, an online meeting management website, to automate critical parts of the invitation and registration processes. With Cvent, we can set up meeting places and times, provide information about location and content, market the events via email, and register participants—online. Updates can be sent to specific registrants or posted to the meeting site.

Cvent permits meeting managers to personalize communication to a target audience, set specific times for delivery of invitations and reminders, and track whether participants are opening and responding to email invitations and reminders. Registration information will be available to both PARCC and Pearson on demand.

The Cvent invitation will include instructions for participants to make transportation arrangements through our travel center. Any airfare or train reservations will be directly billed to Pearson. We will also direct-bill lodging to Pearson and book guest rooms based on participants' registration information in Cvent. A continental breakfast and lunch will be provided for all meeting days.

Participants' car rental (if required), miscellaneous parking tolls, mileage, dinner expenses, and other costs are included in our Cost Proposal worksheets and explained in further detail in the cost narrative.

As previously noted, our Cost Proposal also includes the cost of meeting rooms, supporting equipment (including computers for required meetings), reference materials, office supplies, printing and shipping costs, disposition of secure materials at the conclusion of the meeting, and travel costs for contractor staff to attend the meetings. In addition to the reviewers and contractor staff, we have included travel costs for PARCC staff to attend the meetings as shown in the figure in the previous section and in the Cost Proposal's "Travel and Meetings" tab. Participants can expect to receive reimbursement within four to six weeks after the meeting.

If requested, we can enable solicitation of feedback from your participants after the meeting. By using the Cvent registration software, Pearson can survey participants on such factors as the registration process, advance communications, the meeting facility, presenter, content, and meals.

Once participants have responded, we can provide PARCC with reports on meeting evaluations on either an individual or aggregate level.

Requirement

V.A.1.I. Test Development Specifications and Review Meeting Reports

Response Requirements for Section V.A.1.I.

a) Offeror's proposal shall include a response to the requirements specified in Section V.A.1.1.

Deliverables for Section V.A.1.I.

a) Deliver Annual Test Development Specifications
b) Deliver Reviewer Attendance Report
c) Deliver 8-Day Report (after each review meeting)
d) Deliver 30-Day Report (after each review meeting)

Response

Test Development Specifications and Review Meeting Reports

Within 45 after the contract is executed, or by mid-May, whichever date is later, Pearson will provide our test development specifications to the Partnership. The test development strategy will include at minimum:

- An executive summary
- Key personnel, including their name, role, and contact information
- Key test development milestones
- Test/item development plans, incorporating the requirements within the RFP
- The annual development targets
- Text/passage and item/task targets for each meeting
- A review meeting plan meeting the requirements within the RFP
- The review schedule

The test development specifications will be updated annually in November at least six weeks in advance of development starting for subsequent years.

After each review, Pearson will provide the Partnership with the reports outlined below. Most of these reports can be automated through our item banking system and our online meeting management system. We will work with the Partnership on the exact format, and to verify that the information outlined below is presented in a clear and concise manner.

Eight-Day Report

Within eight business days after reconciliation for each meeting, Pearson will provide the Partnership with reports of the review meeting. These reports will show by grade level and item type for math, and grade level, task, and item type for ELA the number of items Pearson brought to the meeting to be reviewed, the number and percent of any items/tasks not reviewed, number and percent of items/tasks accepted without any edits, the number and percent of items/tasks accepted with minor revisions, the number and percent of items/tasks the committee instructed Pearson to revise and resubmit, and the number of items/tasks the committee rejected. For the text/passage review meetings we will provide the information for text/passages.

Participant Attendance Report

Pearson will provide the Partnership an attendance report no later than twenty business days after each review meeting. This report will be structured for each committee (review group) and broken down by state. The report will be cumulative; after contract award we will work with the Partnership to determine if the cumulating will be by development cycle, calendar year, or something else and the exact format. The report will reflect the number of meetings each individual was invited to, the number of meeting they attended, and the cumulative percent of the meetings they attended.

30-Day Summary Report

Pearson will provide more detailed committee reports no later than 30 business days after the conclusion of each meeting, referred to as the 30-Day Summary report. These 30-Day reports will include item/passage detail and the participant evaluation results.

The item/passage detail will include the following:

- The results for each item, as well as the final outcome/recommendation make for each item or passage
- Running summary that shows the number of math items accepted, including the functionality of technology-enhanced items/tasks
- Running summary results the show the number of completed (full passage sets) ELA passages and item sets to understand the passage has a complete set of items as

defined by PARCC and ready to be field tested, or if additional items/tasks are needed to complete the passage set.

- Graphs that show the summary information by grade and content area

The participation evaluation results will include the results of the evaluations forms completed by the meeting participants for each question. The questions will, at a minimum, include the following:

- If the overview of the purpose of the meeting was clear
- If the participant felt they understood the purpose of the meeting
- If the training on Pearson's item banking system was clear and adequate to complete their assigned tasks
- If Pearson's item banking system was user-friendly
- If the pre-meeting instructions provided by email were clear
- If the group facilitator provided effective leadership for the review
- If sufficient time was provided to complete their assigned task
- Suggestions for other opportunities to assess the spirit of the Common Core Standards

Pearson will review these reports with the Partnership to affect future item development and the meeting planning, meeting documents, and meeting training.

Requirement

V.A.1.J. Data Review Committees

Response Requirements for Section V.A.1.J.

a) Offeror's proposal shall include a response to the requirements specified in Section V.A.1.J.

Response

After items are field tested, Pearson and our administration partners will convene data review meetings involving a committee of educators and citizens from Partnership states. The purpose of the data review meetings will be to identify items that are eligible for operational test construction, items that should be modified and field tested again, and items that should be eliminated. As required, the data review committees will consist of state content experts, grade-level teachers, accessibility and accommodations experts, psychometricians, and higher education faculty, as appropriate.

Pearson will configure the data review committees as indicated in Table V.A.1.J.1 in the RFP. Contractor responsibilities will be shared by Pearson, ETS, and WestEd based on their proposed responsibilities within the contract. For example,

- For ELA, Pearson will provide content experts for grade band 3-5, WestEd will provide content experts for grade band 6-8, and ETS will provide content experts for high school.

- For Math, Pearson will provide content experts for grade band 3-5, and ETS will provide content experts for grade band 6-8 and high school.
- In keeping with their role leading psychometrics, ETS will provide three psychometricians to float among groups. Pearson will send also one psychometrician to the data review meetings.
- Measured Progress, as part of their role performing an independent audit and analysis of the psychometric services, may also send an observer to the data review meetings as appropriate.

As stated in the RFP, the data review committees will be convened in mid-July, after field testing. These in-person meetings will comply with the general test development review meeting requirements as outlined in Section V.A.1.G.3.1 of the request for proposals.

Pearson and ETS will work together to produce the summary materials needed for the data review committee meetings. ETS will conduct psychometric analyses of the field test data as part of their responsibilities on the project. These analyses are described in detail in our response to Section V.C.1 of the request for proposals. ETS will provide any additional psychometric item analysis data as appropriate to answer the questions posed on page 75 of the request for proposals.

The statistics resulting from these analyses will be uploaded to the PARCC item bank and to any additional systems to be used to generate electronic item cards and related materials for the meeting. Pearson will be responsible for generating these materials for the data review meetings. Reviewers will have computers and be able to view items and the corresponding data for each item, as well as interact with the item.

The results of the item review committee meetings will include the final disposition of each item and task evaluated (e.g., move to the operational pool, revise and pretest again, eliminate). This information will be recorded by the content experts and input into the PARCC item bank. Pearson will also provide independent documentation of the data review results to the Partnership chairs for subsequent confirmation.

Requirement

V.A.1.K. Data Review Committee Participants' Travel Costs

Response Requirements for Section V.A.1.K.

a) Offeror's proposal shall include a response to the requirements specified in Section VA1.K.

b) Offerors must include budgeted amounts in their Price Response/Reply and Budget Worksheets to cover the all reimbursable costs for all participants' air travel, car rental, lodging, and miscellaneous expenses such as parking, tolls, vicinity mileage, and other costs Offerors identify as necessary to participants' travel. Participants will also be reimbursed for one meal allowance and/or partial per diem.

c) \Offerors will include budgeted amounts in their Price Response/Reply and Budget Worksheets to cover all reimbursable costs, including but not limited to costs of meeting rooms, and all supporting materials such as audio/ visual equipment and meeting supplies such as flip charts, reference materials, note pads, pens, pencils, and other office supplies, as well as costs related to shipping and/ or producing materials onsite, the disposition of secure materials at the conclusion of the meeting, and other costs Offerors identify as necessary to successfully conduct the meetings.

Deliverables for Section V.A.1.K

a) Deliver Data Review Meeting Plan

b) Deliver Data Review Meeting Agenda, Participant List, Meeting Note

c) Deliver Data Review Training Materials

Response

The meeting and travel requirements in this section are the same as those in section V.A.1.G.3.1. As requested in the RFP, only the differences are acknowledged below.

The first data review meeting will be convened in July 2015, after the first operational year. Pearson will cover all travel and meeting costs for participants, including airfare or rental car, shuttle service to and from the hotel, lodging, meeting room, meeting material, meeting meals, and per diem for dinner and other expenses as described in the Cost Proposal. In addition the $150 daily stipend amount will be provided per meeting day for up to 60 participants, as required in Appendix U.

We will provide a meeting plan to PARCC at least eight weeks before the annual data review meeting. It will include the meeting agenda and participant list for Contractor staff. PARCC will have an opportunity review all training materials prior to the meeting. An updated participant list with state educators and PARCC state representatives will be provided to PARCC closer to the time of the meeting.

Requirement

V.A.1.L. Proposed Item and Passage Review Schedule

Response Requirements for Section V.A.1.L.

a) Offeror's proposal shall include a response to the requirements specified in Section V.A.1.L.

b) Offeror shall recommend the duration of each of the meetings list above.

c) If the Offeror determines that the number of planned meetings is insufficient to review the Partnership-determined number of test items to be reviewed by these committees, the Offeror shall propose creative solutions that allow for committee reviews of all required test items during an annual development period. Solutions may include supplemental virtual and/or in-person reviews within the development year, but the supplemental reviews must adhere to the meeting guidelines outlined in this RFP.

Deliverables for Section V.A.1.L

a) Deliver Review Schedule

Response

Based on the item and passage development plan included in this proposal, Pearson has modified the review schedule accordingly, as shown in the attached and its subcontractors will follow the proposed meeting schedule design as reflected in the updated meeting . It should be sufficient to cover the content review needs of the Partnership for both items and texts. Pearson proposes a combination of in-person and virtual meetings to best meet the needs of the program, but also to be sensitive to meeting costs which can be rather high for in-person meetings. The following meeting schedule is proposed by Pearson to reflect the later start of item development. The reviews to support the development in years 3 and 4 are expected to follow the pattern for year 2.

Meeting	Development year/cycle	Month	# blueprint sets	Descriptions	Mode	% year develop ment	FT year
1	Year 1	Sep-14	2	Text review	in-person	100%	2016
2	Year 1	Jan-15	2 (HS 3)	CLG	in-person	100%	2016
3	Year 1	Feb-15	2 (HS 3)	SE	in-person	20%	2016
4	Year 1	Mar-15	2 (HS 3)	Bias	in-person	100%	2016
5	Year 2 -cycle 1	Feb-15	2	Text review	in-person	50%	2016
6	Year 1	Apr-15	2 (HS 3)	Editorial	in-person	10%	2016
7	Year 2 -cycle 1	May-15	2	CLG	in-person	50%	2016
8	Year 2 -cycle 1	Jun-15	2	SE	in-person	50%	2016
9	Year 2 -cycle 1	Jul-15	2	Bias	in-person	50%	2016
10	Year 2 -cycle 1	Jul-15	2	Editorial	in-person	5%	2016
11	Year 2 -cycle 2	Aug-15	2 (HS 3/int 1)	Text review	virtual	50+%	2017
12	Year 2 -cycle 2	Dec-15	2 (HS 3/int 1)	CLG	virtual	50+%	2017
13	Year 2 -cycle 2	Jan-16	2 (HS 3/int 1)	SE	virtual	10%	2017
14	Year 2 -cycle 2	Feb-16	2 (HS 3/int 1)	Bias	virtual	50+%	2017
15	Year 3 -cycle 1	Feb-16	2	Text review	in-person	50%	2017
16	Year 2 -cycle 2	Mar-16	2 (HS 3/int 1)	Editorial	virtual	5%	2017

The first meeting of any phase should be in-person, but then additional meetings for that year would be virtual. This process has been successfully implemented in practice currently for PARCC and other Pearson developed programs. Section V.A.1.G discusses the length and type of meeting being proposed in more detail.

Given the large number of meetings included under the proposed contract, Pearson will put together a comprehensive meeting plan at the beginning of each year for PARCC approval that will summarize the location, recruitment needs, and other key details for each meeting to be held during the year. Estimates have been done to cover the costs for the meetings as described.

Requirement

V.A.1.M. Item Bank Management and Access

Response Requirements for Section V.A.1.M.

a) Offeror's proposal shall include a response to the requirements specified in Section V.A.1.M.

b) Offeror shall recommend the duration for each of the meetings list above.

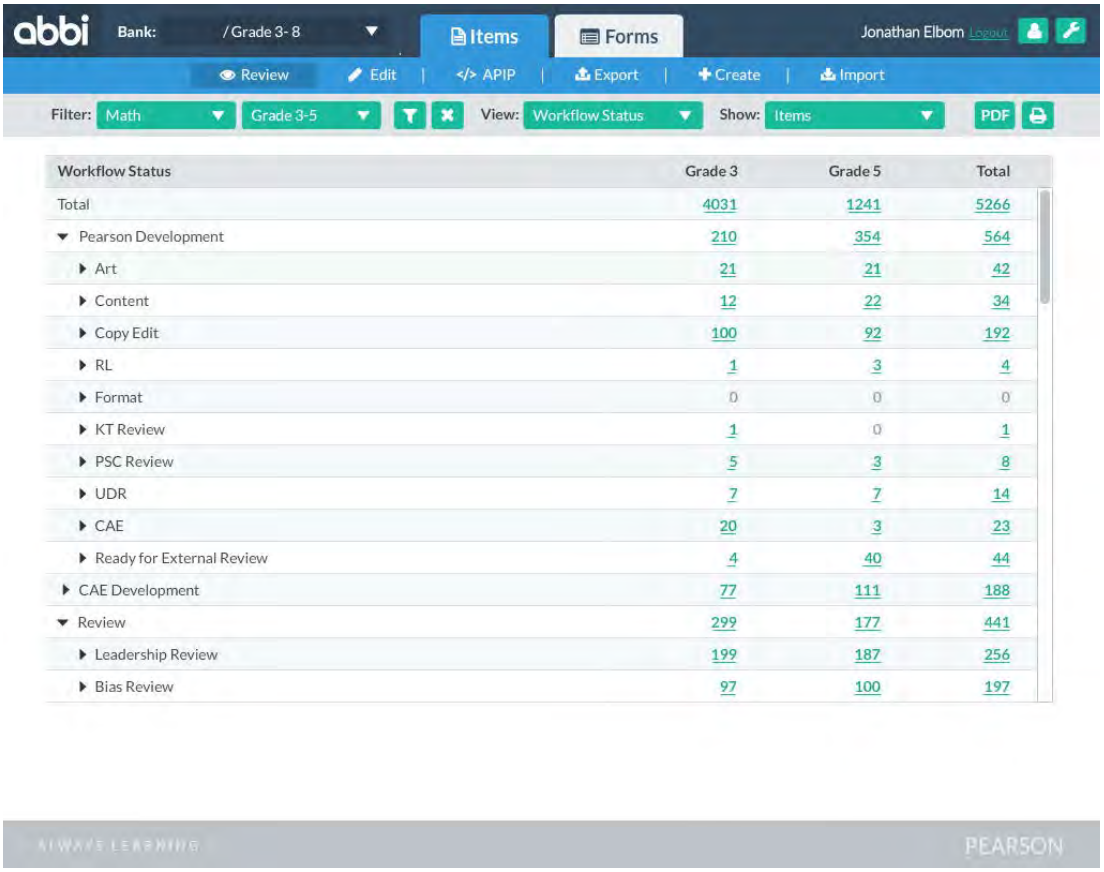
c) If the Offeror determines that the number of planned meetings is insufficient to review the Partnership-determined number of test items to be reviewed by these committees, the Offeror shall propose creative solutions that allow for committee reviews of all required test items during an annual development period. Solutions may include supplemental virtual and/or in-person reviews within the development year, but the supplemental reviews must adhere to the meeting guidelines outlined in this RFP.

Response

Pearson will provide all management, monitoring, and support functions for the PARCC item bank, which includes authoring components for the duration of the contract.

The Pearson item bank provides several user interfaces, data extracts, and reporting functions that will allow PARCC leadership to monitor the health of the item bank and to monitor the content as it moves through the development process. The ABBI content inventory screen provides many views that show item counts across several dimensions including workflow status, learning standard, subject, grade level, item types, etc. Users can quickly assess the status of the content using these interfaces.

The screen illustration below shows how the item inventory can represent item counts by workflow status (expandable options on left side) and grade level. The View option can be used to change the view to other options such as by learning standards or item type. Advanced filtering options are available to locate items by metadata values or other item characteristics. The item counts are hyperlinks so you can view a list of items that are represented by each of the counts.



Workflow Status	Grade 3	Grade 5	Total
Total	4031	1241	5266
Pearson Development	210	354	564
Art	21	21	42
Content	12	22	34
Copy Edit	100	92	192
RL	1	3	4
Format	0	0	0
KT Review	1	0	1
PSC Review	5	3	8
UDR	7	7	14
CAE	20	3	23
Ready for External Review	4	40	44
CAE Development	77	111	188
Review	299	177	441
Leadership Review	199	187	256
Bias Review	97	100	197

Through the security management systems in ABBI, system administrators can provide access to any user by assigning access roles. This would include any PARCC leadership and designated Contractors that need access. There is no limit to the number of users that can be granted access, so supporting the 50 users specified in this solicitation is entirely feasible. Pearson will manage all users access for our content development staff and sub-contracted resources. We will need support from PARCC to manage the PARCC leadership access to add or remove access as needed.

Requirement

V.A.2. Translations

V.A.2.A. Purpose

1. Approach
2. General Test Administration Directions
3. Parent/Guardian Score Reports
4. Mathematics Summative Assessments

Response Requirements for Section V.A.2.

a. Offerors proposals shall include a description of their:

i. Experience in providing translations from English into Spanish and other languages for high-stakes grade K-12 assessments;

ii. Approach to ensuring comparability and validity in translations, including their knowledge of the type and extent of use of dialects (i.e., cultural, social, ethnic, and/or regional varieties), and familiarity with the Common Core State Standards for Mathematics in grades K-12, in order to distinguish between construct-relevant and construct-irrelevant terminology in the standards and aligned assessment items;

iii. Experience providing translation services within a computer-based item exchange system; and

iv. Experience providing audio, video, scripts, computer based and print-based translations.

b. Include in base cost proposal:

i. A plan for translations that describes the steps that will be taken to ensure that all requested translations are accurate and unbiased, and includes a timeline listing the production benchmarks for translations that will ensure translated materials will be available for the first operational year of assessments in 2014-2015.

ii. Translations of general assessment directions into 10 languages

iii. Translations of parent/guardian reports into 10 languages

iv. Translations of PARCC mathematics assessments into Spanish

v, Scoring services for constructed responses to translated versions of the mathematics assessments.

Response

Translations

Pearson will translate the general test administration directions and reporting shells for reporting results to parents and guardians from English into up to 10 languages other than English as needed. We will coordinate the translations for the mathematics items developed by both our staff and ETS. In this response we outline our experience with and approach to translations, and describe our proposed process and schedule for translating the general test administration directions for each assessment, and the reporting shells used to report student results to parents and guardians. In addition, we provide a specific plan and timeline for making translated materials available for the spring 2015 administration and beyond.

Approach

Pearson has extensive experience in developing translations for a variety of our large-scale assessment customers, including Florida, Virginia, Texas, and Illinois. Our recent translation work for New York and Kentucky has involved translating mathematics assessments that have been developed to measure the common core state standards. Our experience includes providing audio, video, and scripts for both computer based and print-based translations. We have experience working through a previous version of our TestNav system with understanding how Spanish characters are typed into open response, and how Spanish XML is represented.

Pearson generally sub-contracts for translation services for large-scale assessment items and materials, and we would propose to do so for the PARCC translation work as well. We partner with a variety of organizations for such translations. For several of our state assessment programs we use Tri-Lin Integrated Services, Inc. Tri-Lin specializes in translation and adaptation of assessment items from English into Spanish. As a subcontractor of Pearson, Tri-Lin researches terminology as well as cultural and regional differences to generate the proper translations of both mathematics and science items. In addition to Tri-Lin, Pearson has utilized Teneo Linguistics, Eriksen Translations, Inc., Betmar Languages, Precision Translating Services, and 1-800-Translate, among others, for various large-scale assessment language translations. Between our assessment and instructional groups, Pearson utilizes a dozen or more companies to provide translation services for our large scale assessments and instructional materials.

Pearson utilizes best practices within the assessment industry for our assessment translation work. This often involves a combination of directly translating a test form's English text to the second language and adapting the content to account for the linguistic and cultural differences between speakers of the two different languages. The combination of translation and adaptation required to produce a reliable and valid version of an assessment in a language other than English is sometimes referred to as *transadaptation*. A transadaptation allows for the translator to make certain changes to adapt the text to the students' native culture. For example, a football could be changed to a soccer ball; a bagel or a hotdog could be changed to something more commonly eaten in Spanish-speaking countries, and so forth. Stansfield (2003) points out that that adaptation may go so far as to involve removing some items and replacing them with others that are more appropriate for the native language or more valid for the examinee population or for the language of the new test. In general, the need for adaptation is lessened through the application of universal design principles in test development. For example, Pearson universal design checklists specifically call out language appropriateness ("The item avoids words or phrases that are sexist, racist, or otherwise offensive, inappropriate, or negative to any subgroup. Language should be simple and clear."), ethnic stereotypes ("The item avoids unnecessary references to and uses the proper reference for ethnic, racial, or cultural groups."), and cultural familiarity ("The item does not rely on an assumed shared experience that is class oriented or native English speaking oriented. Presentations of cultural or ethnic differences should neither explicitly nor implicitly rely on stereotypes nor make moral judgments."). Pearson will follow this process that includes the combination of translation and adaptation (transadaptation) when creating the Spanish versions of the PARCC mathematics assessments.

Several of Pearson's assessment clients require the practice of back translation. A back translation procedure involves translating the source version of the assessment materials (generally in English) into versions for the target language(s), then translating back to, and comparing them with, the source language to identify possible discrepancies. This technique is relatively effective for detecting mistranslation or major interpretation problems and was recommended in early guidance on test translation (c.f., Hambleton, 1994; 2002). However, recent work with international assessments such as PISA and TIMMS has suggested that back translation has a potential deficiency because if a passage is too literally transposed, there is a fairly high risk that the back translation will merely recover the original text without revealing the error (Grisay, 2003). Stansfield (2003) discussed several problems in relying on back translation to examine the quality of the translated document, and Solano-Flores (2012) pronounced back translation as “discredited”. Rather than depending on back translation to check the quality of an initial translation, Solano-Flores (2012) suggested the use of multidisciplinary teams of professionals who review the translators’ translations and, depending on the form of translation accommodation, may also participate in different stages of the translation process.

In keeping with this approach, Pearson recommends the use of teams that would include translators, bilingual educators, test development experts, and additional resources as appropriate (e.g., sociolinguists). This approach involves multiple translations of the same source material and a reconciliation process to verify the equivalence of the translated items or auxiliary test materials to the original English sources.

General Test Administration Directions

Pearson will provide direct translation of test administration directions for ELA/literacy and mathematics content area assessments. Translations will be provided for up to ten languages to be identified by PARCC in the summer of 2014. The 2011 US Census data from 2011 and 2007 shows a pretty stable trend in the languages other than English that are spoken in the home. The information below is the top ten languages for the 2011 census along with the percentage of respondents who said this is the predominant language spoken if not English. Translations are not required for the Fall/Winter 2014 Block but will be required for all subsequent administrations.

Spanish	62.0
Chinese	4.8
Tagalog	2.6
Vietnamese	2.3
French	1.9
Korean	1.9
German	1.8
Arabic	1.6
Russian	1.5
French Creole	1.2

The test administration directions that will be translated from within the manual will be approximately two pages in length. The translated versions will be saved as PDFs to PARCC resource websites including PearsonAccess. Hard copies of the translated instructions will not be printed and shipped to schools or districts. In addition to the translations for the designated languages, Pearson will provide guidelines to test administrators about how directions may be translated for the cases in which translation of directions are not provided.

Parent/Guardian Score Reports

Pearson will translate parent/guardian reports for up to ten languages designated by PARCC in the summer of 2014. We will prepare .pdf versions of each translated report and will provide the versions for posting on PARCC's website and in the PARCC Partnership Resource Center

Mathematics Summative Assessments

Included in the contract, Pearson will translate two forms into Spanish in year 1, and one form each subsequent year, for a total of five Spanish forms per grade/course during the contract period. Forms will be available online and paper (total of 5 online and 5 paper forms to be developed). The online forms will include TEI items. Spanish versions of the assessment are not required for the Fall/Winter 2014 Block but will be available for all subsequent administrations. States needing to provide the mathematics assessment in other languages besides English and Spanish will work with Pearson to negotiate a separate agreement for any additional translations required.

Pearson will provide direct translations of test items, answer choices, and performance-based tasks. For these translations, Pearson proposes to utilize the multidisciplinary approach discussed above. Text translations will be provided for both computer-delivered and paper-delivered forms. In addition, Pearson will provide written scripts for oral translations for the paper-and-pencil administrations. A separate Test Administrator Manual translated into Spanish is not included.

For computer administrations, Pearson will provide translated text-to-speech and/or audio recordings that can be embedded in the computer-delivered platform. Pearson will also provide side-by-side paper forms and translated on-screen version stacked with the English version.

Pearson will provide scoring services for constructed responses written in Spanish. Drawing on our extensive distributed scoring scorer pool of over 100,000 qualified applicants, we anticipate we will be able to hire a sufficient number of Spanish speaking scorers.

Plan and Timeline for Translations

The Project Plan included in the **Other Supporting Material** section includes the major translation activities and milestones for translation of Test Administration Directions, Reporting Shells, and PBA Math and EOY Math assessments. The translation activities/milestones include:

- Actual translations English to other languages
- QA and verification of equivalence

- Posting to PARCC's website and Partnership Resource Center (Test Administration Directions and Reporting hells
- Printing test booklets and answer documents (P/P PBA and EOY)
- Publishing (online PBA and EOY)
- Quality assurance

References

Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*, *20*(2), 225-240.

Hambleton, R.K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, *10*, 229 - 244.

Hambleton, R.K. (2002). Adapting achievement tests into multiple languages for international assessments. In Porter, A.C. and Gamoran, A. (Ed.), *Methodological advances in cross national surveys of educational achievement*. Washington DC: National Academy Press.

Solano-Flores, G. (September 2012). *Translation accommodations framework for testing English language learners in mathematics*. Submitted to the Smarter Balanced Assessment Consortium (SBAC).

Stansfield, C. W. (2003). Test translation and adaptation in public education in the USA. *Language Testing*, *20*(2), 189 - 207.

PARCC

Component 3: Psychometric Services

Overview of Psychometric Services

The psychometric services in support of the PARCC operational assessments are a small but critically important component of the project. We have assembled a team combining the strengths of several subcontractors to deliver accurate, timely, and thoughtful psychometric services.

ETS will serve as the psychometric lead for the contract and have primary responsibility for the deliverables associated with the data analysis of summative, field test, and retest administrations as described in Section V.C.1 of the RFP. As part of their lead role, ETS will be responsible for the psychometric data analysis systems to be used for the project, as described in the Systems for Data Analysis Section. ETS will lead several of the research studies identified in Section V.C.5 and will collaborate with Pearson on two other studies. ETS will also be responsible for the technical documentation for the assessment administration (Section V.C.6).

As primary contractor, Pearson will provide oversight and coordination on the psychometric work. These tasks will include replicating some of the critical psychometric analyses associated with equating and field-test item calibration. Pearson will also be responsible for data forensics, as well as the technology and data requirements outlined in Section V.C.4.

Measured Progress will provide the independent audit and analysis of the psychometric services as required in Section V.C.7. The primary focus for Measured Progress will be the critical psychometric analyses described in the Data Analysis of Summative, Field Test, and Retest Section, but they will also review other relevant psychometric services associated with Component 3. Finally, Caveon Test Security will support Pearson's data forensics analyses by monitoring the internet (and social media) for breaches of test security.

ETS can thoroughly address PARCC's operational psychometric needs given their comprehensive psychometric work during the field test phase. Deliverables include, but are not limited to, the following:

- Design of a comprehensive plan for sampling to verify appropriate representation across the PARCC states
- Development of blueprints, specifications, and linking plans to support the design of the field test assessments
- Design and programming needed to create automated test assembly tools for ELA/literacy and mathematics.

Additionally, ETS is preparing technical memoranda to describe in detail the analyses to be conducted, once field test data are available. The memoranda will summarize the procedures (and the results) used by ETS to, evaluate the performance of field test items, establish the PARCC scales, and conduct special studies requested by PARCC.

ETS is pleased to provide the same commitment and attention to detail throughout the operational phase of PARCC assessments as we have demonstrated during the field test phase. We recognize and appreciate PARCC's groundbreaking work to create next-generation assessments that will provide tools to identify student preparedness for success in college and in the workplace, potential gaps in student proficiency, and the means by which to address those gaps. ETS will be flexible and accommodating as this significant effort evolves; we are fully committed to supporting PARCC and look forward to the opportunity to apply our psychometric expertise to the operational phase of the PARCC assessments.

We offer some of ETS's best and brightest members, including Dr. Lora Monfils, Dr. Venessa Manna, and Dr. Hyeon-joo Oh, among others—to continue our work on the technical underpinnings of the PARCC assessments. At ETS, each psychometrician is supported by an extensive network of nationally and internationally renowned measurement and research scientists. In addition to Dr. John Mazzeo, who directs the Statistical Analysis, Data Analysis and Psychometric Services Group, this network includes Drs. Randy Bennett, Brent Bridgeman, Tim Davey, Michael Kane, Cara Laitusis, Bob Mislevy, David Williamson, and Wendy Yen.

Our staff provides both standard psychometric analyses to support ongoing programs as well as development and implementation of cutting-edge psychometric processes to support innovative assessments.

We support the technical standards described in the AERA/APA/NCME Standards for Educational and Psychological Testing and creates our own stringent set of policies to which our entire organization must adhere: the ETS Standards for Quality and Fairness. Internal audits assess the compliance of each of its programs against these standards. Our staff has expertise and experience in a broad range of areas including: assessment design, measurement models and applications, data analysis and technology research, statistical theory and practice, college-readiness assessments, and international assessments.

V.C.1. Data Analysis of Summative, Field Test, and Retest

The important work that PARCC is now doing and plans to accomplish in the coming years is unprecedented. PARCC summative assessments will inform student progression toward college and career readiness, and will influence the rigor of instruction designed to help students achieve that goal.

We understand the necessity for the PARCC summative assessments to provide comparable results across member states, both within and across administrations and years (Section V.C.D.1 and V.C.5.B.1). The assessments will be performance-based, will be innovative, will rely more heavily on technology (Section V.C.5.B.2), and will be designed to measure the full range of student performance. The assessments will be vertically scaled (Section V.C.D.1), allowing for both status and growth scores (Section V.C.D.1.6); issues regarding accommodations are part of the process throughout (V.C.5.B.3 and 4); and the assessments will allow for international comparisons (Section V.C.5.B.5).

The PARCC goals are ambitious and require a complex, well-defined analysis plan to support successful implementation of the assessments. This plan is outlined in detail in the pages that follow.

We begin with a robust data collection design that draws on the strengths of both the common-item and randomly equivalent groups' designs for calibration and setting the vertical scale. The quality of the data is paramount to the initiation of a new testing program, particularly one of this scope. Of equal importance is a deep understanding of the goals of the assessment and the flexibility and creativity to tackle issues as they arise. Our plan for the design and analysis of the operational assessments builds on the foundational work done during the field test phase.

The PARCC assessments include performance-based (PBA) and end-of-year (EOY) components. We understand the requirement that students' scores on the two components be combined to produce an overall summative score each year, for ELA/literacy and for mathematics. There is an added degree of complexity for ELA/literacy due to the intent to report scale scores for writing and reading.

The PBA tasks have items that will contribute to both the writing and reading scores, whereas the EOY tasks will contribute only to the reading score. Methods to combine the PBA and EOY scores must take into account the implications for within-grade and cross-grade scaling. An important consideration for scaling is the dimensionality within and across the PBA and EOY components, and whether the two components can be fitted with unidimensional IRT models, both individually and when combined into a single summative test score.

As stated in the RFP, it is expected that all PARCC summative assessments will be calibrated and scaled using unidimensional Item Response Theory (IRT). The IRT framework is ideally suited to the PARCC assessments and corresponding measurement goals. In particular, mixed item-format tests consisting of the combined scaling of dichotomous (selected-response), short answer, and performance tasks (Yen & Ferrara, 1997; Fitzpatrick et al., 2006) can be readily accommodated by IRT models, as can vertical scaling. With years of successful application in K–12 testing programs, IRT models have the flexibility and strength to support the PARCC assessments.

The monumental task of establishing a vertical scale, equating test scores in a high stakes context for tests with a large performance assessment component, measuring growth, and maintaining comparability among states and within and across years cannot be understated. We are aware of many of the challenges involved with this next generation of K–12 assessments, and have designed the analyses to provide a sound psychometric solution to best meet these challenges and to best support PARCC goals.

All of the following components of data analyses will be completed every year for each of the three testing phases: operational, field test, and retests (administered during regular testing windows). We will work with the Partnership to verify that appropriate data review occurs following each component.

Requirement

V.C.1.A. Data Cleaning

Response

Typical Rules for Data Cleaning

Our psychometric team realizes the critical importance of using valid data for analyses. We will determine exclusion rules for analysis in collaboration with PARCC, their technical advisors, and Pearson, and clean the data accordingly. These rules will include removal of duplicate students, removal of students with invalid response strings, and removal of students who did not meet attemptedness rules.

After receiving data from Pearson, ETS will examine the match of the dataset to the data map to verify that the responses for variables, particularly item and total scores, in both the paper-based and computer-based forms are within expected ranges. We will examine out-of range values, and we will consult with PARCC to determine how best to implement the predetermined rules. We will document data cleaning decisions.

Further, ETS will review item responses to each test form and test section to confirm that actual item responses in the data file match the responses expected for that section. For example, if a group of students is assigned to take a specific test form, item responses should only include responses to items in the corresponding test form.

Pearson psychometric staff will be in communication with ETS as data analysis extracts are evaluated and cleaned for analysis. Pearson will facilitate any needed dialogue with Pearson administration processing staff to address questions or resolve any data anomalies.

The following six rules are typical to verify that the data used for analyses are based on a valid set of student responses.

1. Valid Attempt Rule

Students are treated as having a valid attempt if they have answered at least some pre-prescribed number of items at the beginning of the test (for example, the first 5 questions) or at least some pre-prescribed total number of items on the test. We will exclude student responses not meeting the minimum criteria for attemptedness from analyses.

2. Valid Response Time Rule

For computer-based tests, we can also use timing information to determine whether a student has a valid attempt on the test. For example, students may need to spend a minimum amount of time per item. Timing information can also be used to judge student motivation level.

For example, a student's item score may be removed if the student completed the test in a very short time and received a very low score. Item response time can also be used to indicate the possibility of interruptions in the test delivery/administration. Negative values, zeros, and extremely large values will be flagged as these values.

If item response time is equal to zero or smaller, we may classify the value as missing. If item response time is extremely large (exceeding the time limit), the corresponding item scores may also be classified as missing. Decisions on how to use item timing data will be made jointly with PARCC.

3. Form Validation Rule

Given that multiple test forms will be developed for PARCC, ETS will verify that students were administered the appropriate form for each specified test administration. If the student's form code does not match the forms specified for a particular administration, we will flag the score so that if PARCC requests, it may be excluded from item analysis and calibration.

4. Records Invalidated by the Test Administrator

An invalid test record may include one for which there was an allegation of cheating reported for a specific test administration, administration of an inappropriate form (e.g., seventh graders were given the eighth grade test), or some other unexpected occurrence (e.g., interrupted testing because of school emergency). Typically, invalidated test records are not included in item analysis and calibration.

5. Missing Data

Omitted responses occur for a number of reasons. When a page or screen layout is complicated or following a long passage or task, students may not respond due to confusion or fatigue. Missing data can also occur when tests are speeded and students run out of time. Other reasons for omitted responses include instances where students are unmotivated, extremely anxious, or overwhelmed.

When students omit responses to items, the items typically are scored as zero or not reached. Determining how to best deal with omitted responses requires considering the underlying reason(s) for the omissions and the uses of the data. We will work with the PARCC to determine the optimal treatment of missing responses.

6. Missing Score Levels

Another case of unplanned missing data can occur with performance tasks (constructed response) in which score levels are missing in the collected data—that is, when no student has achieved a certain score level. This gap in scores most commonly occurs at the highest score level.

While this is an unplanned source of missing data, it is something that can occur in new testing programs or when introducing new item types. This type of missing data may indicate a problem with the item, scoring rubric or simply indicate that the item is extremely difficult and/or measures knowledge and skills unfamiliar to the student.

In this case, the general recommendation, depending upon the source of the problem, is to update the rubric and collapse score levels; eliminate the item from the pool; and/or refine the item in a subsequent development phase and re-administer the item in the future. However, we recognize that there may be times when the item is deemed appropriate for inclusion on an operational test form for content and/or policy reasons.

Data Cleaning Rules for Various Testing Phases

Clean data is always critically important. Of course, procedures for attaining clean data may vary depending on the testing phase. Valid attempt, response time, form code, and record invalidation by administrator rules would likely hold regardless of test phase. The treatment of missing data, however, may vary in different testing phases.

For item calibrations during field testing, where the goal is to evaluate item performance, omits can be treated as not reached so that item difficulty estimates are not unduly affected by omitted responses.

In contrast, the treatment of omits for operational scoring is typically a policy decision, because the way in which omitted items are scored can affect the student's ability estimate or score.

We will work with PARCC to determine the business rule for treating omitted responses for all item types for both field test and operational items.

System to Be Used for Data Cleaning

At ETS, we have a unit, the Data Quality Services (DQS) that is dedicated to evaluating and promoting the quality of all data sets delivered to our Psychometrics group. The work that DQS performs is a critical part of ETS's data quality control processes.

ETS psychometricians and DQS staff will create a PARCC-specific, automated validation program that will run predefined checks on all data files and will verify that all fields and data needed to perform the statistical analyses are present and within expected ranges. This program will be SAS-based, in keeping with the goal of using open-source or commercially available software for all analyses.

Analyses to Identify Valid Response Data

Several analyses are available to identify valid response data. We will use descriptive statistics and frequency tables to check variables in the dataset for their expected ranges, averages and variability. Graphical representation of the data may be used to check for outliers.

Further, ETS will validate students' demographic information on the form against the file provided by Pearson. We will verify that all students are accounted for in the data file, and eliminate duplicate files. We will compare the item and total scores on the form with the scores in the file to verify a match for each student.

Working with Various PARCC States

We will work to create final data cleaning rules that are acceptable to PARCC. For example, accommodation rules may vary by state. Therefore, we will be sensitive to these variations to create rules for how to treat responses on special forms such as braille and large print.

ETS has years of experience working with states to confirm that their specific accommodation needs are supported in K–12 state assessment programs, this includes both testing and reporting requirements. Implementation of these rules will be contingent upon PARCC's approval.

Requirement

V.C.1.B. Classical Item and Test Analysis

Response

Classical Item Analysis

We will perform classical item and test analyses for all phases of the testing program. Classical item analyses involve computing a set of statistics for every item in each form on the test. Each statistic is designed to provide some key information about the quality of each item from an empirical perspective. It is also a quality control step to verify answer keys and that the item is performing as expected for the purpose of contributing to student scores. The information is used for item reviews, test construction, revisions, technical reports, and other psychometric analyses and documentation.

After receiving all of the student response data, implementing scoring rules, checking the data files and applying agreed-upon valid case criteria to the data, the next step will include a classical item analysis to evaluate item difficulty, item discrimination, and student raw score performance of selected-response items, hand-scored and machine-scored constructed-response items. These analyses are used to identify any items that might not perform as expected. We will evaluate the following classical item statistics and will work with Pearson to determine the standard criteria for item difficulty, item-total correlation, distractor-total correlation, percentage of items omitting an item, score point distribution, and time on task.

Item Difficulty (*p*-values)

For dichotomously scored items (i.e., selected-response items), this statistic indicates the proportion of students in the sample who answered the item correctly. Desired *p*-values generally fall within the range of 0.25 to 0.95. For polytomously scored items (i.e., constructed-response items), this statistic represents the average item score or the proportion of the maximum obtainable score.

Desired values generally fall within the range of 30 to 80 percent of the maximum obtainable score. Items that fall outside the desired difficulty range may not have performed as expected due to a number of factors such as lack of familiarity with the item type or opportunity to learn.

Item-Total Correlation

Selected-Response and Constructed-Response Items. This statistic describes the relationship between performance on the specific item and performance on the entire test form. It is sometimes referred to as a discrimination index. For selected-response items, the item-total correlation is the point-biserial correlation and for constructed-response items, the item-total correlation is the polyserial correlation.

Typically, values of 0.20 or higher indicate a desired positive relationship. We will flag items with lower item-total correlations and our psychometric staff, in collaboration with the assessment development team, will carefully review these items.

Items with negative correlations may indicate a scoring key error or serious problem with the item content, for example, multiple correct distractors, or confusing presentation of question to be answered.

Distractor-Total Correlation

Selected-Response Items. This statistic describes the relationship between selecting an incorrect response for a specific item and performance on the entire form. Typically, the magnitude of the correlation between an incorrect answer and test (or form) performance is expected to be negative. The values of this correlation are typically compared and contrasted with the discrimination index.

When the magnitude of these item-total correlations for incorrect answers is stronger than for the correct answer, we will carefully review the item for content-related problems and for the possible revisions prior to re-field testing.

Percentage of Students Choosing Each Response Option

Selected-Response Items. This statistic indicates the percentage of students who selected each of the answer options and the percentage that omitted the item. Fully functioning answer options are essential to each item. An item option that is selected by few if any students indicates a problem with the plausibility of the option.

Percentage of Students Omitting an Item

Selected-Response and Constructed-Response Items. This statistic is useful for identifying problems with test features such as testing time and item/test layout. Typically, it is assumed that if students have an adequate amount of testing time, approximately 95 percent of students should attempt to answer each question. When a pattern of omit percentages exceeds 5 percent for a series of items at the end of a timed section, this may indicate that there was insufficient time for students to complete the items.

For individual items, commonly used flagging criteria include an omit percentage greater than 5 percent for a single selected-response item or 15 percent for a constructed-response item (omit rates are typically greater for constructed-response items than for selected-response items, therefore a higher omit percentage is typically used to flag a potential constructed-response item/test layout problem). High omit rates may indicate there is an item/test layout problem; for example, students might accidentally skip an item that follows a lengthy stem. Such items should be considered for revision and may need to be field tested again.

Score Point Distribution

Constructed-Response Items. Investigation of the distribution of scores is helpful in identifying how well an item is functioning. It also provides empirical justification for the scoring rubrics. If very few or no students are assigned to a certain score point, this may indicate that the item is not functioning as expected, that the scoring rubric is flawed in some way, that there are problems with the item content, and/or that students have not been taught the content. In this case, the item needs to be flagged for further action. The rubric may need to be adjusted or the item may need to be revised for future field testing.

Time on Task

We will also summarize time on task for each item. Tracking time on task for each item is particularly useful in the test assembly process since knowing how long it students require on average to takes to respond to a set of items allows us to control for speededness. Items that take a long time to answer compared to other similar items can be flagged during item analysis as potentially problematic.

Psychometric Analysis Specifications

Upon contract execution, ETS psychometricians will create psychometric specifications for all proposed analyses activities. The analysis specifications will include flagging criteria. Items and tasks can be flagged as poorly functioning based on one or more of the classical item analysis statistics described above. Pre-defined flagging criteria allow the efficient and consistent identification of items/tasks for further examination.

We will provide the draft specifications to PARCC and work with PARCC and Pearson to revise until approved. We suggest that the following item-flagging rules, along with suggested values be considered for use in the review of the statistics:

- Items with *p*-values above or below a specified threshold (e.g., above 0.95 or below 0.25)
- Item-total correlations below a specified threshold (e.g., 0.20)
- Distractor-total correlations above a specified threshold (e.g., 0.00)
- Greater number of high-performing students choose a distractor than the keyed response
- High percentage of omits (e.g., greater than 5 percent)
- High percentage that do not reach the item (e.g., greater than 5 percent)
- For constructed-response items, items with a low percentage of students obtaining a score point (e.g., less than 5 percent)

We will conduct item analysis for both operational (OP) and field test (FT) items. For OP items, we will remove flawed items from scoring, with PARCC's approval. For FT items, we will revise or review for further action any flawed items with the PARCC's approval.

Test Analysis

In addition to the item statistics described above, we will also complete test analysis. For example, for each operational form, we will summarize the mean and standard deviation of raw and scaled test scores along with the number of students taking each form, the average *p*-value and the average point biserial. We will also demonstrate and report test reliability for the student population as a whole and subgroups of interest, including by state. To show the reliability of the student scale scores and the proficiency level assignments, we also will report the conditional standard error of measurement for each scale score point and the classification consistencies for each proficiency level.

Once cut scores have been established, we will summarize the percentage of students in each performance level for all students and by subgroups of interest. Regarding the equating of forms, we will compare the mean, standard deviation, and raw score distributions of the anchor set for the reference and new groups, as well as evaluate the items using the criteria listed above to identify potentially flawed items. See section V.C.1.D.for more information on evaluating equating results.

Our team is committed to working with PARCC to verify that the classical item and test analysis conducted for this contract are completed following best practices and in a way that supports the goals of PARCC.

Requirement

V.C.1.C. Differential item functioning for field test, summative assessments, and retest assessments

Mantel-Haenszel Procedures

One of the goals of test development is to assemble a set of items that provides an estimate of a student's ability that is as fair and accurate as possible for all groups within the tested population. Differential item functioning (DIF) statistics are used to identify those items that identifiable subgroups of students (e.g., males, females; white, Asian) with the same underlying level of ability have different probabilities of answering correctly.

If the item is differentially more difficult for an identifiable subgroup when conditioned on ability, the item may be measuring something different from the intended construct. However, it is important to recognize that DIF-flagged item performance differences might be related to actual differences in relevant knowledge or skills (item impact) or statistical Type I error.

As a result, DIF statistics are used to identify potential sources of item bias. Subsequent review by content experts and bias/sensitivity committees are required to determine the source and meaning of performance differences.

Given the innovative nature of the PARCC items, the goal of measurement along the full range of ability for diverse learners, and the relatively small sample sizes, two observed score methods are recommended. The Mantel-Haenszel (MH) procedures (Dorans & Holland, 1993; Mantel & Haenszel, 1959) have been used widely to provide a summary measure of differential functioning for the comparison groups as a whole. The logistic regression (LR) method of DIF detection (Swaminathan & Rogers, 1990) is recommended to provide insight into differential functioning at various points in the ability range.

DIF analyses will be conducted for designated comparison groups of interest, assuming sufficient sample size. These may include groups based on demographics (gender, race/ethnicity, economic disadvantage), special instructional needs (SWD, ELL), or other factors of interest (for example, jurisdictions or states). DIF analyses are typically not conducted if the sample size for either the reference group or focal group is less than 100 or the sample size for the two groups combined is less than 400.

Selected-Response Items

The Mantel-Haenszel (MH) statistic (Dorans & Holland, 1993; Mantel & Haenszel, 1959) will be used for selected-response items. This statistic is expressed as the differences between members of the "focal group" (for example, female, Asian, African American, Hispanic, and Native American) and members of the "reference group" (for example, males and White) after conditioning on ability (for example total test score).
The MH procedure is one of the more commonly used methods to detect DIF. This method uses contingency tables to compare the probability of success on each item for the studied groups of interest after matching on overall ability (i.e., total test score). The common odds ratio is estimated across all categories of matched examinee ability. The resulting estimate is

interpreted as the relative likelihood of success on a particular item for members of two groups when matched on ability. As such, the common odds ratio provides an estimated effect size where a value of unity indicates equal odds, and thus no DIF (Dorans & Holland, 1993).

The odds ratio takes on values from 0 to infinity and is interpreted as the average factor by which the odds that an examinee of the reference group will answer an item correctly exceed that of a member of the comparable focal group. Values less than unity indicate DIF in favor of the focal group, a value of unity indicates the null condition, and a value greater than one indicates DIF in favor of the reference group. The associated MHχ2 is distributed as a chi-square random variable with 1 degree of freedom.

As an index of magnitude, the odds ratio is frequently transformed to a delta scale given by MH D-DIF = -2.35 in ($\hat{\alpha}_{MH}$) where positive values indicate DIF in favor the reference group and negative values favor the focal group.

Constructed-Response Items

DIF analyses of the constructed-response items will be completed using two procedures. The first is the Mantel-Haenszel (MH) ordinal procedure. Based on the Mantel procedure (Mantel, 1963; Mantel & Haenszel, 1959), it compares the proportions of matched examinees from each group in each polytomous item-response category—that is, the probability of a given item score for the studied groups of interest after matching on total test score.

As with dichotomously scored items, the common odds ratio is estimated across all categories of matched examinee ability. The resulting estimate is interpreted as the relative likelihood of a given item score for members of two groups when matched on ability.

The associated Mantel chi-square statistic is used in conjunction with a second procedure which is the standardization procedure (Dorans & Schmitt, 1993). This procedure produces a DIF statistic based on the standardized mean difference (SMD) in average item scores between members of two groups who have been matched on their overall test score. The SMD compares the item means of the two studied groups after adjusting for differences in the distribution of members across the values of the matching variable (total test score). A negative SMD value means that, conditional on the matching variable, the focal group has a lower mean item score than the reference group.

In contrast, a positive SMD value means that, conditional on the matching variable, the reference group has a lower mean item score than the focal group. The SMD is divided by the standard deviation (SD) of the total group item score in its original metric to produce an effect-size measure of differential performance.

Classification

Based on the DIF statistics and significance tests, items are classified into one of three categories and assigned values of A, B, or C. Category A items contain negligible DIF, Category B items exhibit slight to moderate DIF, and Category C items have moderate to large values of DIF. Negative values imply that, conditional on the matching variable, the focal group has a lower mean item score than the reference group. In contrast, a positive value implies that, conditional on total test score; the reference group has lower mean item score than the focal group.

The following figures show the flagging criteria for selected-response and constructed-response items.

DIF Categories for Selected-Response Items	
DIF Category	**Criteria**
A (negligible)	Absolute value of the MH D-DIF is not significantly different from zero, or is less than one.
B (slight to moderate)	1. Absolute value of the MH D-DIF is significantly different from zero but not from one, and is at least one; OR 2. Absolute value of the MH D-DIF is significantly different from one, but is less than 1.5. Positive values are classified as “B+” and negative values as “B-”.
C (moderate to large)	Absolute value of the MH D-DIF is significantly different from one, and is at least 1.5. Positive values are classified as “C+” and negative values as “C-.”

DIF Categories for Constructed-Response Items	
DIF Category	**Criteria**
A (negligible)	Mantel Chi-square p-value >0.05 and |SMD/SD| ≤ 0.17
B (slight to moderate)	Mantel Chi-square p-value <0.05 and |SMD/SD| >0.17
C (moderate to large)	Mantel Chi-square p-value <0.05 and |SMD/SD| > 0.25

Logistic Regression Procedure

In recent years, logistic regression (Agresti, 2002; Howell, 2010) has been proposed as a model-based approach for identifying uniform and non-uniform DIF (Swaminathan & Rogers, 1990). The advantages of using logistic regression instead of the MH statistic for DIF identification include the use of continuous variables as independent variables in the logistic regression equation, the possibility to model uniform and/or non-uniform DIF, and the possibility of generalization of the binary logistic regression model for use with ordinal item scores (Zumbo, 1999, p. 24).

For selected-response items scored as incorrect or correct, the item response (0 or 1) is used as the outcome measure in the regression equation, and the logistic curves for the groups are constructed (Zumbo, 1999). Although logistic regression is less popular in practice than the Mantel-Haenszel (MH) statistic, the two methods together will be complimentary and should maximize the detection of DIF. If the logistic regression curves corresponding to the two groups are parallel, this indicates that there is no interaction between the group membership and the ability level. If these two curves are separate from each other, the uniform DIF (the systematic advantage for one group over the other exists across all ability levels) exists.

Alternatively, if the logistic curves for the two groups are not parallel, it is an indication of the interaction between the group membership and the ability level. In this case, non-uniform DIF exists – where the probability of answering the item correctly is different for the two groups across various ability levels (Atar, 2006, p. 9).

Requirement

V.C.1.D. Calibration, Scaling, and Equating of summative assessments

1. Establishing the PARCC Scale
2. Special Consideration: High School Mathematics Assessments
3. Equating Across Years
4. Evaluation of Linking Items
5. Claim and Sub-Claim Level Scores: Raw Scores and Domain Scores
6. Growth Scores

Response

Establishing the PARCC Scale

As specified in the RFP and discussed earlier in the proposal, the two major components of the PARCC summative assessment are the PBA and the EOY. The PBA will be administered several weeks before the EOY. Student performance on the two components must be combined to provide reliable scores for classification into one of five performance categories and to determine year-to-year change or growth in performance.

The purpose of the calibration, scaling, and equating analyses is to establish reporting scales, and to produce scores that are comparable within and across administrations. Scale scores are used to support score interpretations. To support comparisons across levels (such as grades), scores from tests at different levels are placed onto to a common scale through implementation of vertical scaling procedures. To support comparisons within level (or grade), equating procedures are used to maintain score comparability within and across administrations.

Because of the importance of accurate calibration, scaling, and equating analyses to the PARCC assessment program, ETS will be supported by Pearson psychometric staff who will replicate critical IRT analyses related to the generation of reported scores. These replications will include targeted item calibration, linking, and scaling analyses and will be identified as part of the statistical analysis procedures developed prior to each administration. Note that these analyses will be independent of additional quality control audits carried out by Measured Progress as part of their audit and analysis of the psychometric services (see Section V.C.7).

We understand that PARCC intends to report scale scores for reading, writing, overall ELA/literacy, and mathematics. Where feasible, we will vertically scale these scores to support measurement of growth. We will determine performance standards relative to the ELA/literacy and mathematics scales. We will establish all four of the reporting scales (reading, writing, ELA/literacy, and mathematics) using census data following the year 1 operational administration and we will maintain these scales by equating across years.

In the following sections, we begin with a description of the data collection design and then describe the analyses we will conduct to develop the operational scale. As stated in the RFP, it is expected that the PARCC summative assessments will be calibrated using unidimensional IRT models. In using IRT to develop the operational scale, we will build on the dimensionality, model selection, and vertical scaling investigation conducted during the field test phase to determine the extent to which vertical scales are feasible for PARCC assessments. If feasible, we will develop vertical scales at the master claim (scale score) level or at further granulated levels based on learning progressions across grades. .

Data Collection Design

As noted in the RFP, PARCC is planning to offer three administration windows—the Traditional Year window for grades 3-8 and high school courses and the Fall/Winter Block and Spring Block for high school courses. , Online forms, Spanish translations of the mathematics tests, and Integrated Math tests will not be available for the Fall/Winter 2014 Block but will be available for subsequent administrations.

PARCC assessments are designed to be delivered via dual mode (computer, paper), with the goal to move to fully online delivery. Within the next three years, PARCC intends to have 97 percent of students test online, and the proportion of TEI items expanding to 50 percent of the assessment.

Forms for the mathematics PBA and EOY, and the ELA/literacy EOY will consist of Core Forms, comprised of items that contribute to the operational score, plus matrix sampled items that do not contribute to the operational score. Matrix sampled items will include embedded field test items and items used for linking (horizontal and/or vertical). Forms for the ELA/literacy PBA will not include matrix sampled items.

It is assumed that for the first operational year, there will be five Blueprint sets that will yield 10 Core forms by repeating items/tasks from a given Blueprint set on more than one Core form. .

Given the parameters above, the design for the PARCC operational data collection must address competing pressures associated with costs, practicality, and the expected quality of results. In general, within the context of an operational assessment program, one seeks designs that minimize item exposure and test burden without unduly affecting the quality of results. The most effective data collection designs are robust to common sources of errors but remain practical to implement.

The data collection design must incorporate some means of linking together, onto a common scale, the items and tasks administered to different groups of students. Two methods of linking are commonly used:

Common Items. This requires that blocks of items and tasks be administered to different groups of students; depending on the intended scaling purpose, these blocks may reflect on-grade (or “within-grade”) content or off-grade (or “cross-grade”) content. To obtain within-grade (horizontal) scaling, common items reflect the grade-level blueprint. For cross-grade or vertical scaling, the common items specify the articulation in content across grade levels

Randomly Equivalent Groups. In this approach, the test content is randomly assigned to separate samples of students from the same population. The test material presented to different student samples is considered as comparably “on scale” by virtue of the equivalence of the groups.

Each approach has advantages and disadvantages. For example, the common item approach is dependent on the common items performing equivalently across groups, however, item position and context effects may prevent this. Regardless, common items are also capable of providing strong linking bonds. In contrast, the equivalent groups method is efficient but vulnerable to lack of random equivalence between groups.

We propose an integrated approach to the data collection design that has the advantages of both linking methods to place the operational item parameter estimates onto a common scale. Moreover, this approach addresses the need to minimize item exposure across the assessment windows.

The result is a design that is both reasonably efficient and robust enough to defend against many potential sources of error. The data obtained are ideally structured for item response theory (IRT) calibrations. The designs also incorporate common-item links between grade levels to establish a vertical scale. These links are implemented by administering blocks of test content sampled from the adjacent lower or upper grade level at most grade levels.

We will scrutinize content administered from an upper grade to a lower grade to minimize concerns regarding students' opportunity to learn.

Test Design for Common Item Linking

Common item linking is accomplished by using both the overlap in forms created by re-use of items/tasks from given blueprint sets, as well as the matrix sampled items—specifically the following:

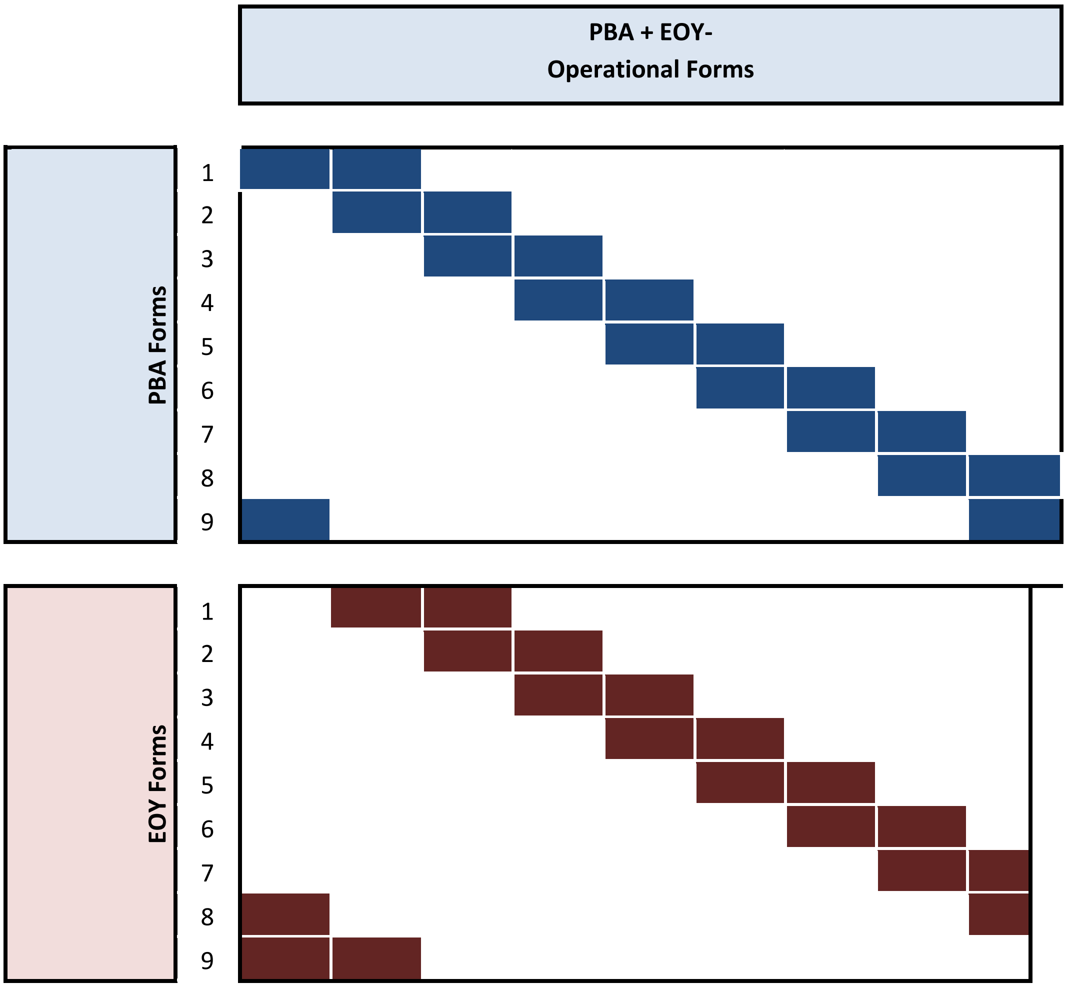
- Within grade, certain pairs of forms will overlap partially with one another by containing some proportion of Core Form items in common (depending on the length of the forms and the nature of the items). To the extent it is supported by the test blueprint, we would propose a modular approach (20-30 percent targeted overlap). This will provide common item linkage for placing IRT parameter estimates for all items within grade on a common scale.
- Paper and online forms will share items for the purpose of linking and comparability studies.
- Per the RFP, mathematics PBA and EOY forms and the ELA/literacy EOY forms will have a variable set of matrix-sampled items (matrix section) that is external to the Core Forms. The matrix section will be populated with either embedded field test items or items for cross-grade linking. In addition, the matrix section can be used to supplement the on-grade common item linkage. Within the matrix sections, each student is administered a small number of items.
  - For the purpose of vertical linking, the matrix sections will be populated with off-grade items from the grade above and the grade below. The off-grade items will appear as on-grade items in Core Form for their corresponding grades (e.g., grade 4 items will appear as off-grade items in grade 3 and grade 5 forms and as on-grade items in grade 4 forms).
  - We have assumed the number of matrix sections needed to support PARCC's stated field test requirements; we assume matrix sections can be randomly assigned to students in the case of computer-based administration or spiraled among students in a classroom for paper-based administration.
  - Form construction specifications will include criteria for selection of matrix-sampled items to support needs of comparable scores, for establishment and maintenance of vertical scales, and for replenishing the item bank through field testing.

  - Criteria for selection of matrix-sampled items for individual forms will take into account the appropriateness of the mix of content and item-types relative to the set of operational items, and the impact on the testing experience of the individual test taker.
  - Strategies for distribution of matrix sampled items across forms, whether for linking (horizontal or vertical) or for field-testing, and distribution across students will take into account the number to be matrix-sampled as well as the need to minimize individual testing time and minimize security risks through item over-exposure.
- In contrast, the ELA PBA forms will not have any external matrix sections. These forms will have partial overlap of Core Form items/tasks between pairs of forms that will serve as on-grade linking items.
  - Currently, the PARCC form design for the ELA PBA does not support the administration of off-grade tasks/items for the purpose of vertical linking. Administered items are limited to on-grade operational Core.
  - It may be possible to use the standalone field test for the ELA PBA that is immediately adjacent to the EOY operational administration. To do this, we would spiral in some number of PBA tasks from the current year's operational administration among the field test items/tasks—off-grade items/tasks for vertical scaling, and on grade items/tasks to serve as common item linkage to the PBA administration to adjust for changes in item difficulty due to the later administration in the academic year.

    This approach would be viable only if sufficient sample sizes are available, students are motivated to interact with the tasks in a meaningful way, and there are sufficient tasks available to provide construct coverage for the adjacent grades to be linked.

The essentials of the data collection design for within-grade and vertical scaling are illustrated briefly in the following figures. For ease of interpretation, the number of forms and number of item blocks that comprise a given form is limited so as to better communicate the overlap of items across forms. For a given grade, the partial overlap between certain pairs of forms is shown conceptually in the following figure. For example, within grade, PBA forms 1 and 2 partially overlap, as do EOY forms 1 and 2.

Note that high school mathematics courses requires special treatment. We address these later in this section.



Data Collection Design for a Given Grade

The following figure provides a depiction of the test content of on-grade and off-grade item assignment by grade. The lowest grade (i.e., grade 3) will have items from grade 3 and items from the adjacent upper grade (i.e., grade 4). Similarly, the highest grade (i.e., grade 11) will have items from grade 11 and items from the adjacent lower grade (i.e., grade 10).

For the remaining grades, off-grade items will be drawn from the grade immediately above and the grade immediately below. For example, forms for grade 4 will have items from grade 4, and items from grades 3 and 5. As noted previously, the off-grade items will appear in the matrix section that is external to the Core Form.

Test Content by Grade	Examinees by Grade: 3	4	5	6	7	8	9	10	11
3	**3**	3							
4	4	**4**	4						
5		5	**5**	5					
6			6	**6**	6				
7				7	**7**	7			
8					8	**8**	8		
9						9	**9**	9	
10							10	**10**	10
11								11	**11**

Spiraling Plan to Support Randomly Equivalent Groups

This design also takes into account the potential for a long test administration period to accommodate the academic calendars of the PARCC states and the corresponding need to maintain test security within and across administrations. Depending on the length of the overall test administration period, it may be advisable to partition the administration into two or more windows, and develop a spiraling plan that limits the number of forms available for administration in each window.

Under a distribution plan of this type, there may be overlap in form distribution across windows (within or across years) such that a certain number of forms are designated for large volume distribution in any given window, and one or more versions of those forms for low volume distribution in a subsequent window. To increase stability of results (in the aggregate, and as pertains to horizontal and vertical linking), we recommend a spiral design with a broad range of material to minimize error and bias.

We can easily adapt the hypothetical spiraling plan illustrated below to accommodate the relative volumes and desired number of testing windows (and sub-windows) within an overall administration period, and the number of available test forms. Forms are spiraled within windows, and this provides randomly equivalent groups linkage to supplement the common item linkage for horizontal equating to support test score comparability.

Distribution of Test Forms across Testing Windows

Window	Exposure Level Due to Distribution Volume (Forms)			
1	High (1,2,3), Low (4,5)			
2		High (4,5,6), Low (7,8)		
3			High (7,8,9), Low (9,10)	
…				…

Sample Size. PARCC has indicated that the operational scale will be based on census data. However, the number of students for the matrix-sampled items will be substantially smaller, and will depend on the number of matrix-sampled item sets that are administered within each window.

To support IRT calibrations for the vertical linking items, we recommend that the data collection for each administration mode provide representative samples of at least 1,500 valid cases per item/performance task per form, including sufficient numbers of Students with Disabilities (SWDs) and English Learners (ELs) for special analyses and research studies. These recommendations can be considered in the context of overall sample sizes and the number of matrix sections, but are of particular interest for external field testing such as is being considered for the ELA PBA tasks.

To achieve this target, we propose a minimum overage of approximately 20 percent, or 300 students per item (1,800 total per item), to allow for attrition, non-response, and other factors that may reduce the yield of usable item response data.

System Used to Calibrate and Scale PARCC's Assessments in a Single Grade

Within the family of IRT models two major decisions need to be made: 1) to implement a unidimensional or a multidimensional model approach, and 2) to use a Rasch one-parameter/partial credit model (Rasch/PC) combination, a two-parameter logistic/generalized partial credit model (2PL/GPC) combination, or a three-parameter logistic/generalized partial credit (3PL/GPC) combination.

Under the Field Test Assessment Administration contract, ETS will conduct a dimensionality study as part of a scoring and scaling study to investigate the choice of IRT models. The results of these analyses will help to inform our proposed solution for calibration and scaling of the assessments based on operational data. Pending these results, a variety of scaling and scoring options can be considered.

In this section, however, we will assume that we will employ a unidimensional IRT model for scaling and scoring. If, however, there is strong evidence of multidimensionality, as may be likely in the case of reading and writing in ELA/literacy, we will explore separate unidimensional IRT calibrations for each domain and development of a composite scale for the overall ELA/literacy test.

Operational implementations of IRT models in K–12 testing programs historically have been unidimensional models. We know that student achievement is not affected by only one trait or ability, and there is no pretense that unidimensional models are identifying or measuring only one trait. In these models, the trait or scale on which items and students are jointly ordered is identified, in essence, as the major dimension that best explains student performance.

The Rasch/PC and 3PL/GPC unidimensional models have been extensively demonstrated, in K–12 testing programs, to provide strong horizontal equating (i.e., between test forms within a grade or course) as long as the content and statistical specifications are well maintained for every test form. That is, even in cases where assessments are not unidimensional, a unidimensional model can provide accurate equating as long as the multidimensionality reflected in the content specifications is essentially consistent across test forms (Reckase, Ackerman, & Carlson, 1988).

ETS will carry out all analyses using commercially available and/or open-source software. Various software options are presented in Systems for Data Analysis.

IRT Item Calibration

The usefulness of IRT models is dependent on the extent to which the models effectively represent the data. Based on the results of the field test analyses, if a particular set of models is recommended, we will validate these assumptions or evaluate the assumptions of models under consideration.

In the text that follows we discuss IRT model assumptions, model-data fit, and precision of item estimations and how we will examine these aspects to evaluate the IRT item calibration. We begin with a discussion of dimensionality analyses that we will implement to validate the results of the field test for the operational assessment data.

Dimensionality Analyses

Prior to IRT scaling, we will conduct dimensionality studies within each summative test component (PBA and EOY) and grade, across the PBA and EOY within each grade, and across grades. Dimensionality analyses are informative for determining the use of a unidimensional versus multidimensional model in IRT scaling, the score aggregation method for PBA and EOY tests, and the feasibility and structure of a vertical scale.

In particular, we will conduct the dimensionality analyses as follows:

- Apply tests of essential unidimensionality (i.e., DIMTEST; Stout, 1987) to assess whether the PBA and EOY, individually and when combined, within a grade deviate from the unidimensionality assumption. DIMTEST is a nonparametric, IRT-based approach that examines the conditional covariance between two distinct, homogeneous subsets of items.

  In essence, it tests whether the reference composites for the two subsets of items point in the same direction at different points along the scale. If there is no significant difference between the dimensions measured by these items, the data can be treated as essentially unidimensional. We will use students who complete a PBA or EOY form in the dimensionality analysis of that test, and we will use students who finish both PBA and EOY within a grade in the dimensionality analysis of the summative test (PBA and EOY).

- If the analyses above indicate possible multidimensionality, we plan to examine the dimensional structure of the PBA and EOY tests individually and combined within grade using both exploratory and confirmatory methods. For the exploratory approach we plan to use a parallel analysis and the vector approach developed by Reckase, Martineau, and Kim (2000). The purpose of this examination is to identify the number of dimensions at each grade level.

  For the confirmatory analysis we plan to examine the items using a between-item dimensional structure (i.e., simple structure) where each item is only allowed to load on a single factor. The possible factors to be considered are item type (selected-response, constructed-response, and technology-enhanced items), content strand, form, and/or general factor (as in a bi-factor model). We will use these results to determine the extent to which the data depart from a unidimensional factor structure.

Validating IRT Model Assumptions

Different IRT models have different underlying assumptions. Some common assumptions are related to properties of the latent space (e.g., dimensionality and local item independence), equal discrimination, examinee guessing, and non-speededness of test administration. Below, we describe the key evidence we will use to validate these assumptions, as well as other analyses we will conduct to evaluate the success of the IRT calibrations.

Dimensionality

The dimensionality study and associated methodology were discussed above. We will review the results of the dimensionality investigation and evaluate their implications for IRT model applications. A primary piece of evidence will be the degree to which the PBA and EOY assessments measure the same construct. From a unidimensional perspective, we will evaluate item fit. Since there are many items for any grade/content area combination, we will compare the distributions and plots of item fit.

Equal Discrimination in Rasch/PC

We will review the distribution of classical biserial/polyserial correlations, and item discrimination parameter estimates under 3PL/GPC and 2PL/GPC model calibrations, and their relationships to item difficulty. A reasonably homogeneous distribution indicates overall good model-data fit if the selected model assumes equal item discrimination.

Minimal Guessing in Rasch and 2PL

If Rasch or 2PL is the IRT model of choice, we will conduct an evaluation of the amount of guessing involved for selected-response items by examining the size of guessing parameter estimates under 3PL/GPC model combinations. We will also review the test difficulty, time limits, and item format to assess the possible role of guessing in test performance.

Local Independence/Minimal Testlet Effect

Items that are part of an item set or passage-based are likely to produce local item dependence (LID). We will use two methods, as described below, to check LID for these items:

1. We will conduct Item and testlet reliability analyses (Wainer & Thissen, 2001) to evaluate the degree of LID. Specifically, for each item type that involves context-dependent item sets, we will compute and compare two reliability estimates. The first reliability estimate assumes that all items are locally independent, while the second reliability estimate is calculated with items treated as testlets.

   If the reliability estimates computed using the item-level data are considerably larger than those computed using the testlet data, then there is evidence for LID. Item types that show notable LID may be treated as testlets in IRT calibration, which will provide a more accurate description of the item function lines and the information provided by these items.

2. We will use generalized residuals (Haberman, 2009; Haberman & Sinharay, 2013) for item pairs based on an IRT model to check LID. Specifically, we will concurrently calibrate all test forms within a grade based on a recommended IRT model.

   We will check the generalized residuals for proportions of students in the score categories of item pairs among the items that have a common setting to see whether LID is evident in these items. We can compute the generalized residuals approach by the MIRT package (Haberman, 2009) which was developed at ETS and is free for non-commercial use.

Evaluating Goodness-of-Fit

We can use a likelihood ratio χ_j^2 test statistic to compare the frequencies of correct and incorrect responses in the intervals on the θ continuum with those frequencies expected based on the fitted model (du Toit, 2003):

$$\chi_i^2 = 2\sum_{h=1}^{n_g}\left[r_{ih}\log_e \frac{r_{ih}}{N_h P_i\left(\bar{\theta}_h\right)} + \left(N_h - r_{ih}\right)\log_e \frac{N_h - r_{ih}}{N_h\left(1 - P_i\left(\bar{\theta}_h\right)\right)} \right],$$

where n_g is the total number of intervals, r_{ih} is the observed frequency of correct responses to item *i* in interval *h*, N_h is the number of examinees in interval *h*, $\bar{\theta}_h$ is the average ability of examinees in interval *h*, and $P_i\left(\bar{\theta}_h\right)$ is the value of the fitted response function for item *i* at $\bar{\theta}_h$.

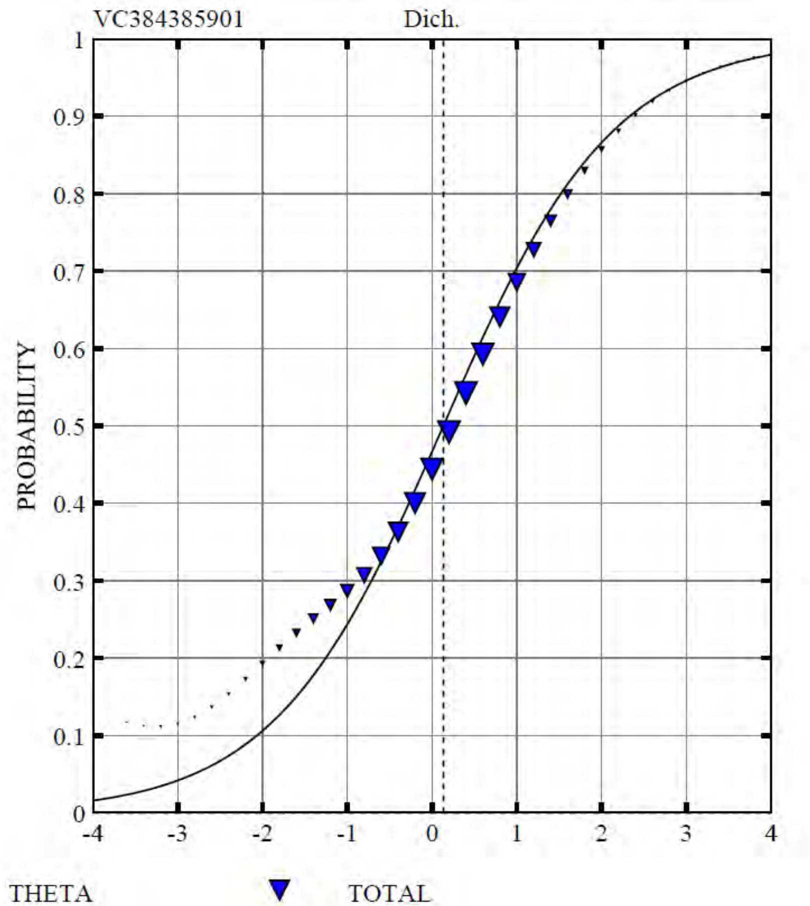
Because the statistic tends to be sensitive to sample size (i.e., flagging more items for large sample sizes), we will develop realistic cut-off points for item misfit flagging. In addition, we will use graphical evaluation in conjunction with the goodness-of-fit statistic.

Graphic Evaluation: Residual Analysis

Residual analyses have proved to be a helpful and effective way to understand and interpret data. First, examinees are classified into a number of ability subgroups. Then, for each subgroup, the observed item performance is plotted and compared to the expected item performance as determined by the item characteristic curves (ICCs) based on the item parameter estimates obtained under different IRT model combinations (Hambleton, Swaminathan & Rogers, 1991).

We will use a computer program called “PARPLOT” (ETS, 2009) to produce item residual plots. The following figures show sample residual plots for a dichotomous item and a polytomous item, respectively. The line represents the expected item performance and the triangles represent the observed item performance with the size of the triangles proportional to student sample size. For an item to show good model data fit, it is expected that the triangles, especially the large-sized triangles, scatter closely around the line.

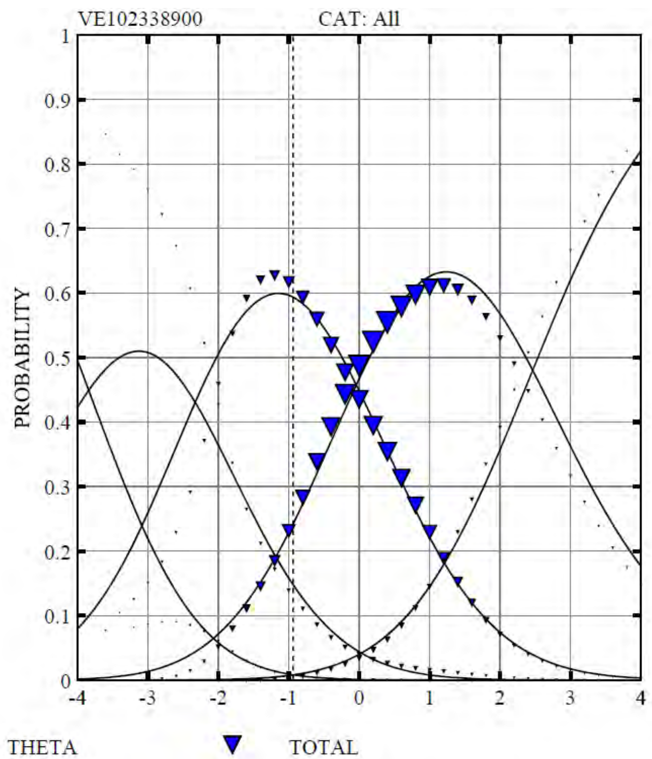
Sample Residual Plot for a Dichotomous Item



Parameters

	VALUE
A	0.588000
B	0.133561
C	0.000000
P+	0.4948

Sample Residual Plot for a Polytomous Item



Parameters

	VALUE
A	0.588000
B	-0.936416
D1	2.903870
D2	1.391498
D3	-0.895568
D4	-3.399800

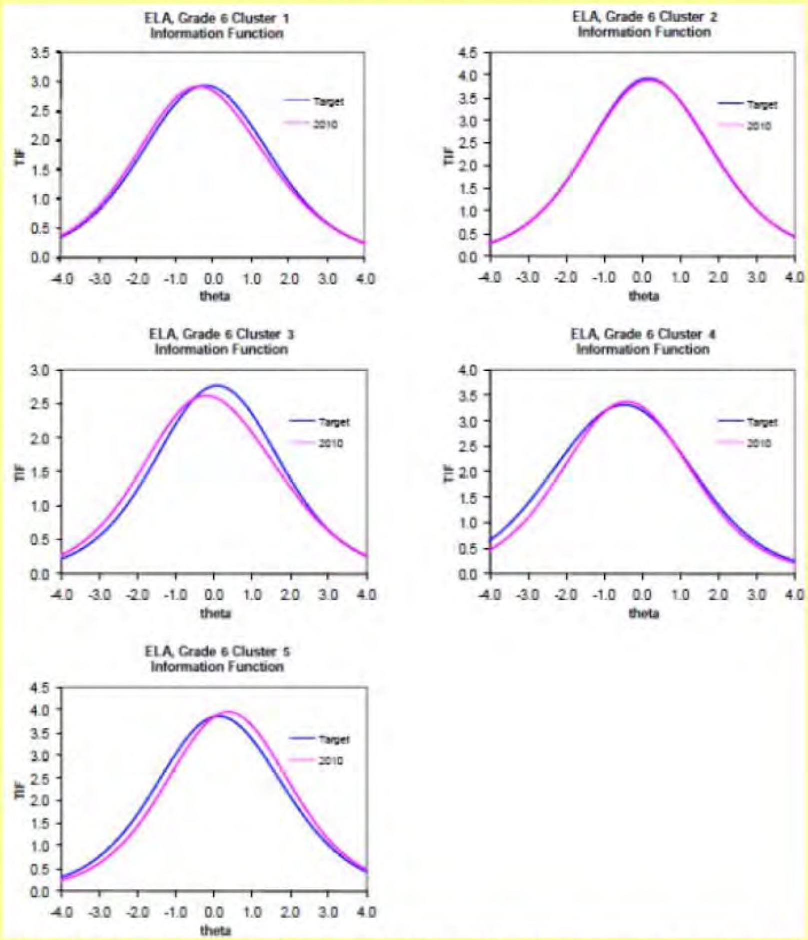
Precision of IRT Estimates

For each test form, we will examine plots of the test information functions (TIFs) and associated conditional standard errors of measurement (CSEM) for total test scores and subscores. These graphics support comparative analyses across administered forms and may be used to identify target curves for future form assembly. Plots of TIFs provide a way to readily compare overall and subscore information for each test against target TIFs and previously administered test forms.

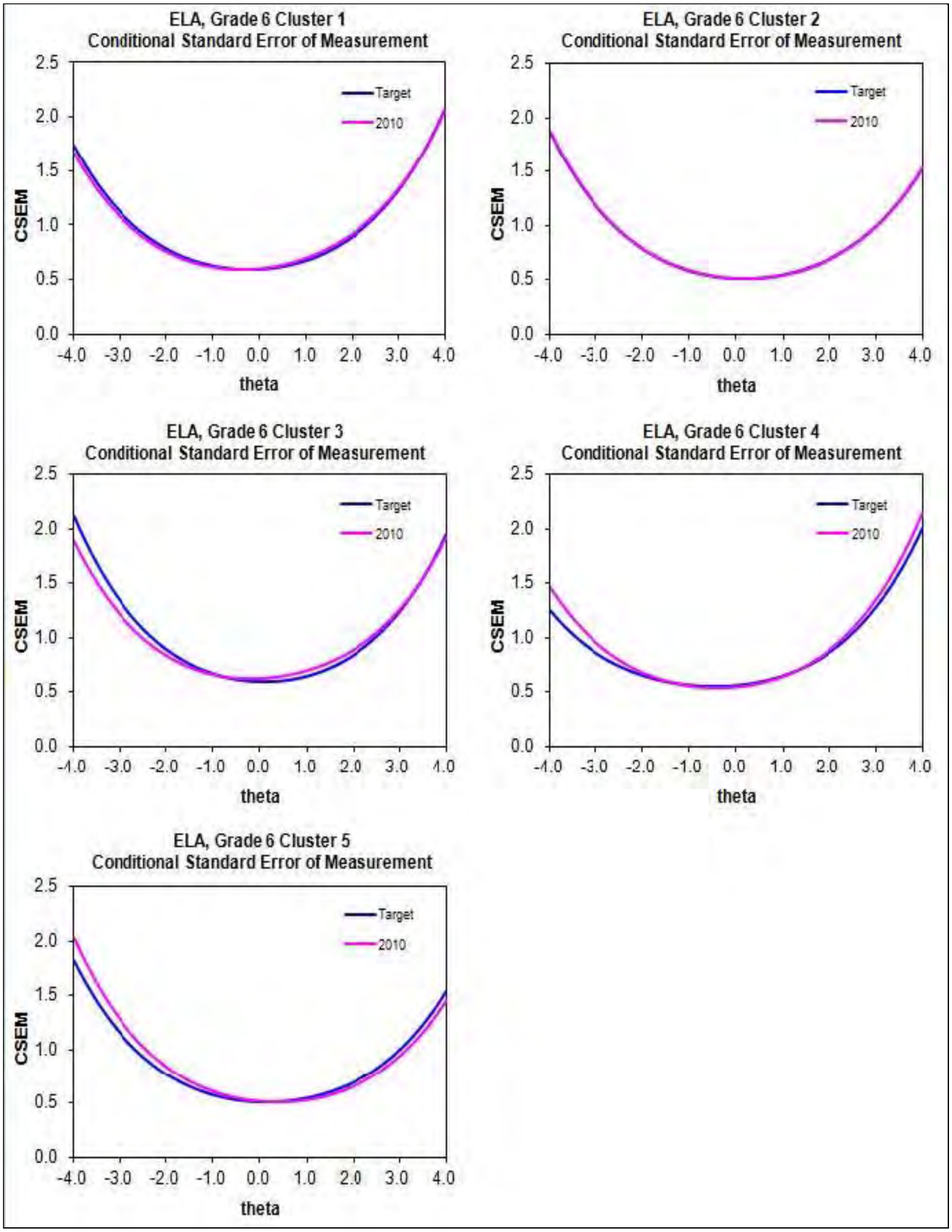
In addition, IRT information may be used to evaluate gaps in the item collection relative to the test blueprint at the level of total test, subscore, or content standard. Comparison of total test and subtest TIF and CSEM plots against targets supports construction of forms that measure with comparable precision across the ability range from one administration to the next. This information will be tabled in the form of raw to scale conversion tables.

We will provide results of these analyses to PARCC and its Technical Advisory Committee (TAC) for review prior to finalizing the on-grade scales (and vertical scales as applicable).

The following figure provides examples of such comparisons below at the level of content cluster. ETS will be responsive to PARCC requests for adjustments to these procedures.



Plots of Subtest Information Functions (TIFs). Plots of TIFs provide a way to easily compare overall and subscore information for each test against target TIFs and other pilot forms.



Plots of Subtest Conditional Standard Error of Measurement (CSEM) Curves. Comparison of total test and subtest TIF and CSEM plots against targets support construction of forms that measure with the comparable precision across the ability range from one administration to the next.

Proposed Scaling Method

The purpose of IRT calibration and scaling for a single grade is to place items within each grade level on a common difficulty scale to the extent supported by the results of dimensionality analyses.

For horizontal scaling, we will conduct methods for simultaneous or concurrent calibration of items at each content area/grade level by means of a hybrid of the common items and randomly equivalent groups linking approaches. The “common items” approach requires blocks of items and tasks that partially overlap be administered to different student samples.

For the “equivalent groups” approach, the test material presented to different student samples is considered as comparably “on scale” by virtue of the equivalence of the groups. Neither of these linking methods is guaranteed to work perfectly in practice; therefore, the linking design we propose incorporates both strategies.

As described in the section on Data Collection Design, this is accomplished by common item linking across test forms and by spiraling forms within test windows. The result is a design that is both reasonably efficient and robust to many potential sources of error and well-structured for IRT calibration.

Method for Combining PBA and EOY Scores

The ELA/literacy and the mathematics tests are both comprised of EOY and PBA components that differ substantially in both content and format. The ELA/literacy test can be further subdivided into reading and writing domains.

The main advantage of having a single scale encompass the entirety of a test is simplicity of interpretation. Item calibration, scale linking (both within and across grade levels) and test scoring are much easier when a single scale is employed.

Use of multiple scales requires separate calibrations and linking for each scale, greatly increasing and complicating the work needed to maintain the test operationally. It can also affect the quality (or at least the efficiency) of item parameter estimation with smaller numbers of items in each subcomponent. .

Methods to combine the PBA and EOY scores must take into account the implications of within-grade and cross-grade scaling. Dimensionality is an important consideration for scaling within and across the PBA and EOY components; this factor affects whether the PBA and EOY components can be both fitted individually with unidimensional IRT models and then combined into a single summative test score (PBA plus EOY).

Assuming that both components can be fitted by unidimensional IRT models, we will consider two score aggregation methods to combine PBA and EOY scores into one summative score:

- Concurrent calibration of both test components to produce one estimate of ability based on responses to both PBA and EOY items/tasks
- A weighted composite of the two IRT ability scores (thetas) from the separate calibrations of the PBA and the EOY, using weights determined by statistical considerations (e.g., reliabilities) or policy considerations

Vertical scaling decisions will influence the choice of a score aggregation method. If a single vertical scale is found to be feasible for the PARCC summative scores (i.e., the PBA and EOY are combined as a single test when creating the vertical scale), the concurrent calibration method for score aggregation is preferred to preserve the properties of the vertical scale.

On the other hand, if separate vertical scales are developed for the PBA and the EOY components, then a weighted composite of separate vertically scaled calibrations would be possible, with weights selected to yield summative scores that reflect consistent and interpretable cross-grade ordinality of each component and the composite.

For a weighted composite to be successful, the number and nature of the items comprising each component must be such that scales are not unduly affected by idiosyncratic item or form effects within and across administrations. This is a potential issue for performance based assessments with small numbers of items and/or forms with novel/innovative item types that may vary in format and/or number from one year to the next.

ETS will use the aggregate scores generated by both methods for students who took the full summative test (PBA plus EOY) to assess the implications for measuring student achievement and growth. For both methods, appropriate weights can be applied to each test to reflect the degree of content importance and coverage in each test. We will work with content experts and test developers to determine appropriate weights, if needed, with the approval of PARCC.

Note, however, that for ELA/literacy, there is an added complexity due to the intention to report scale scores for writing and reading as described below. We will work with PARCC and its TAC to develop an optimal solution to meet PARCC's needs.

Method for Establishing Separate Reading and Writing Scales

We can use the same methods described above to combine the PBA and EOY scores to establish separate reading and writing scales and an ELA/literacy composite scale. If the reading and writing domains are found to represent a single unidimensional scale, we can calibrate both test components concurrently to first establish the ELA/literacy scale.

We can then establish the separate reading and writing scales by using item parameters obtained from each individual component. If the reading and writing domains are found to represent separate or unique unidimensional scales, we can first establish the reading and writing scales, and then use a weighted composite of the two scales to create the composite ELA/literacy scale.

There are potential challenges in calibrating and scaling the PARCC writing test with IRT models. If we cannot use IRT to scale the writing assessment, we can use a classical equating approach such as the equipercentile equating method. If a classical equating approach is required for the writing test, we will establish the ELA/literacy scale as a weighted composite of the reading IRT-based scale and the writing non-IRT-based scale). ETS psychometricians and content experts will work with PARCC and its TAC to determine appropriate weights for each component.

Scoring Table Production

The purpose in providing scale scores is to allow for year-to-year comparisons of test scores for a given content area. For a given test, these scaled scores will be comparable across future forms because any differences between the forms (e.g., mean difficulty) is taken into account during the calibration and equating of item parameters.

To produce scale scores for the PARCC assessments, we propose to use an IRT number-correct scoring procedure (Yen, 1984) to develop raw score to scale score conversion tables. The IRT calibration and equating process provides information in the theta metric (mean 0, standard deviation 1). Because this scale is not particularly useful for reporting purposes, students' raw scores on the operational tests will be converted into scale scores. This is completed by following a two-step procedure: (1) nonlinear monotonic transformations of the raw score points into theta metric points, and (2) linear transformations of the theta points into scale score points.

First, raw scores are mapped to theta score points using a corresponding Test Characteristic Curve (TCC). A TCC is defined as the sum of Item Characteristic Curves (ICC) for the items used for determining student scores for each test form. An ICC represents the probability that a student of a given ability will give the correct answer for that item. In a general sense, an ICC shows the expected scores that students at each given ability are expected to achieve on the particular item. The TCC, as the sum of the ICCs, shows the expected total score that a student of given ability is likely to achieve.

Using the inverse of the TCCs constructed from the item parameter estimates obtained by calibration (and equating, when applicable) of the student response data, each raw score between chance level and the maximum possible score is mapped to a corresponding theta score. The results can be described in a transformation table that converts raw scores to theta score. One conversion table is produced for each test at each grade level.

Second, the theta score is converted to the reported scale score metric via a linear transformation. Thus, through this two-step process, each raw score point is converted to a corresponding theta score which is subsequently converted to a scale core. The general form of the function used to translate the theta points to scale score points is:

$$SS = m1^{*}\ \theta\ + m2$$

- where θ is the theta score corresponding to the raw score point to be transformed;
- *m1* is a multiplicative scalar constant;
- *m2* is an additive location constant; and
- *SS* is the resulting scale score point.

The maximum likelihood procedure cannot produce scale score estimates for students with perfect scores or, depending on the IRT model, scores of zero or scores below the level expected by guessing. Also, while maximum likelihood estimates are available for students with extreme scores other than zero or perfect, occasionally these estimates have standard errors of measurement that are very large, and differences between these extreme values have little meaning. Therefore, scores are established for these students based on a linear interpolation method. These values are called the lowest obtainable scale score (LOSS) and the highest obtainable scale score (HOSS). To assist with interpretation of scores, CSEMs are typically provided for each score point in the metric of the reporting scale.

Selecting Vertical Linking Items

The procedure described in this section makes the assumption that a unidimensional structure across at least some number of adjacent grade levels will be supported by the dimensionality analyses.

The proposed approach to vertical scaling will be separate calibrations by grade with grade-by-grade linking using the Stocking and Lord (1983) test characteristic curve (TCC) technique. In general TCC methods such as Stocking and Lord have advantages when compared to moment methods such as mean/mean or mean sigma (Baker & Al-Karni, 1991; Kolen & Brennan, 2004). When used with separate calibration, the test characteristic curve methods are more robust to violation of the IRT assumptions and reduce error (Hanson & Béguin, 2002).

The item parameter estimates produced as result of separate calibrations at each grade level will be available for adjacent grade linking. We will use the TCC linking method to link the item calibrations at adjacent grades. We will repeat this iterative cross-grade linking process for the adjacent grade pairs, so as to place item parameter estimates on a single (vertical) scale for ELA/literacy and for mathematics, respectively.

We will designate a grade level at the midpoint of grade spans as the base grade for linking. The direction of the linking will be as follows:

$$3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7 \leftarrow 8 \leftarrow 9 \leftarrow 10 \leftarrow 11$$

That is, we propose to use grade 7 as the base test grade for ELA, and link other grades to grade 7 by means of the common item sets shared between adjacent grades. Specifically, we will link grade 8 to grade 7, link grade 9 to grade 8, and so on to grade 11. Likewise, we will link grade 6 to grade 7 and link grade 5 to grade 6, repeating the pattern through to grade 3.

We will apply the TCC linking method to the summative test forms comprised of both PBA and EOY tests. Both ELA/literacy and mathematics are assessed in grades 3 through 11. For ELA/literacy, there are grade-level tests for each grade. For mathematics, there are grade-level specific tests for grades 3-8, and end-of-course tests (EOCs) for high school mathematics that correspond to a sequence of three courses in each of two curricular pathways (Traditional: Algebra I, Geometry, Algebra II; Integrated: Integrated Mathematics I, II, and III). For mathematics, we will employ a multi-faceted approach to evaluating a vertical scale that spans grades 3–11. We will first work with the grade-level tests for grades 3–8. Analyses to investigate the feasibility of including the high school assessments (EOCs) in the vertical scale in mathematics are described in Section V.C.1.D.2.

Cross-Grade Linking Item Refinement

The linking process is iterative and involves an inspection of differences between the new grade transformed item response functions and the reference grade item response functions for each of the linking items. We will flag linking items that show large differences between reference and transformed new response functions and route them for review by assessment development specialists.

PARCC will review any items we recommend for removal from the cross-grade linking set. If approved, we will eliminate these items from the cross-grade liking set and re-run the Stocking and Lord scaling with the transformation parameters re-estimated.

To summarize the difference between new and reference item response functions, we will calculate root mean squared deviation (RMSD) and mean absolute deviations (MAD) for both uniform and weighted ability distributions:

$$RMSD_i = \sqrt{\frac{\sum_{j=1}^{n_g}\left[P_{Ni}^{*}\left(\theta_j\right) - P_{Ri}\left(\theta_j\right)\right]^2 f_j}{\sum_{j=1}^{n_g} f_j}},$$

$$MAD_i = \frac{\sum_{j=1}^{n_g} \left| P_{Ni}^{*}\left(\theta_j\right) - P_{Ri}\left(\theta_j\right) \right| f_j}{\sum_{j=1}^{n_g} f_j},$$

where θ_j is the mean of the abilities in the ability interval *j*, n_g is the number of intervals, f_j is the frequency for ability interval *j*, $P_{Ni}^{*}\left(\theta_j\right)$ is the expected score for item *i* at ability level θ_j based on the transformed new item parameter estimates, and $P_{Ri}\left(\theta_j\right)$ is the expected score for item *i* at ability level θ_j based on the reference item parameter estimates.

To calculate RMSD and MAD based on the weighted ability distribution, we will use the combined ability distribution from the new transformed group and the reference group. To calculate unweighted RMSD and MAD, we will use the uniform ability distribution. We will further evaluate linking items that meet the criteria below and may need to eliminate these items from the cross-grade linking set. However, we may need to relax these criteria somewhat in light of the phase of PARCC development. We will tabulate the number of cross-grade linking items that are flagged based on these criteria.

- Biserial/polyserial value < 0.20 based on either the reference or the new ability distribution (the removal of linking items based on this criterion can be conducted before vertical scaling)
- Weighted and unweighted WRMSD > 0.125 for dichotomous items and > 0.15 for polytomous items
- Weighted and unweighted MAD > 0.15 for dichotomous items and > 0.20 for polytomous items
- In general, cross-grade linking items are expected to show student growth from a lower grade level to a higher grade level. Assessment development specialists will carefully review items that show a reverse pattern (i.e., items that have lower average item score at a higher grade level) before we consider them for removal from the cross-grade linking set.

In addition to evaluating individual linking item performance, we inspect correlations between the new and reference item parameter estimates. In general, high correlations are expected. This correlation tends to be slightly higher for mathematics than for ELA/literacy, possibly because ELA/literacy items are passage-dependent and more susceptible to context and position effects.

Linking Design for Establishing the Vertical Scale

As described previously, the data collection design for establishing the vertical scale draws upon the strengths of both common item and randomly equivalent groups linking. The two basic approaches to vertical linking consist of separate and concurrent calibrations that are used in accordance with other techniques (e.g., choice of a scaling model).

Separate Calibration Approach

In the separate calibration approach, we estimate the item parameters at each grade level. We arbitrarily select one grade level as the base for the scale. Taking advantage of the linking items, we then use linear transformation methods (such as Stocking and Lord, 1983) to place the estimates onto the same scale as the base grade level.

We repeat this process across grades in a chain until the vertical scale is complete. The methods used for separate calibration are not fundamentally different from those used in horizontal equating except that the resulting vertical linkage is evaluated holistically across grade levels.

Concurrent Calibration Approach

The other linking approach for vertical scaling is called concurrent calibration. Concurrent calibration is a multigroup (non-equivalent) method that estimates underlying population distributions (mean and standard deviation) for each group (Mislevy, 1987; Bock & Zimowski, 1997).

This method calibrates students and grade levels in a single step that theoretically uses all the available information. This process results in a large, sparse data matrix since items that were not administered at a particular grade level to students are designated as not reached. The not-reached item designation results in those responses not being included in the likelihood function.

Advantages and Disadvantages of Separate and Concurrent Calibrations

Separate and concurrent calibrations each have their respective strengths and weaknesses. Many studies that have investigated separate versus concurrent calibrations are inconclusive, are limited in some respects, or found no substantive differences (Kim & Cohen, 1998; Hanson & Beguin, 2002; Ito, Sykes & Yao, 2008).

Some modest degree of model misspecification is expected in most applications. With the concurrent approach there is a single step in which the parameters are estimated with no comparable linking error. An important issue is the extent to which the separate or concurrent estimation procedures perform when there is partial model misspecification. When violations of unidimensionality exist, separate calibration may be preferred to mitigate these effects (Kolen & Brennan, 2004).

An advantage for the separate calibration approach is that having two sets of item parameter estimates can help identify and remediate potential problems that may arise. For example, if an item functioned in a very unstable fashion across grade levels and a cause for this lack of stability could be identified (e.g., a large item-position change), then this item could be removed as a cross-grade linking (common) item. These types of problems will be more difficult to detect in the case of a concurrent calibration in which a single set of item parameters is calculated.

Given the innovative nature of the PARCC assessments, the potential for multidimensionality, and the possibility for linking items to perform differently across grade levels, we recommend separate calibrations followed by adjacent grade linking. However, we know that there is a study to investigate the feasibility of vertical scaling as part of the Field Test Assessment Administration contract. The results of this vertical scaling analysis will help to inform our proposed solution to develop and evaluate a vertical scale based on operational data.

Number of Linking Items

An appropriate set of cross-grade linking items is required to establish the PARCC vertical scale, and as such it requires careful consideration from both content and psychometric perspectives. Given the importance of the development and maintenance of the vertical scale, we propose that the cross-grade linking items be selected first, before developing operational test forms.

The optimal number of items in a robust cross-grade linking set is typically 20 percent to 25 percent of the operational test length or no fewer than 20 items, whichever is greater. These linking sets would consist of items that met the minimum criteria for acceptance (as defined above), cover the full range of difficulty, and are as representative of the item types and test content as possible. As noted previously, the cross-grade linking items will be in the matrix section of the operational test forms.

To provide for greater representation of the off-grade content, PARCC may wish to consider use of all appropriate items from one or more full blueprint sets from each adjacent grades. Under this option, the sets of vertical linking items would be spiraled across matrix sections of the operational forms.

Cross-Grade Linking Items

Unlike a horizontal common-item linking set, cross-grade linking items will not span the difficulty or sample the content of the two respective grades. Rather the cross-grade linking items define the expected overlap in the construct across adjacent grades. In particular, the set of linking items for vertical scaling should reflect the growth continuum as it pertains to advances from one grade to the next in the learning progressions across grades toward college and career readiness, as articulated in the CCSS. Moreover, linking items should be selected for instructional sensitivity, as judged by content experts and educators.

Psychometricians, content specialists, and test developers will work together and, as applicable, use field test results to identify items in appropriate content strands as vertical linking items in operational test forms.

We will pay careful attention to establish that content experts identify standards that are applicable to both the on-grade and cross-grade samples of students. In addition, these linking items should meet the psychometric requirements for common item linking blocks. These criteria typically include:

- *p*-values/IRT *b*-parameter estimates
- item-test correlations/IRT *a*-parameter estimates (if applicable)
- omit rates
- DIF
- item fit

After thorough internal reviews and approvals, we will submit the proposed linking sets to PARCC and its TAC for review and approval.

Impact of Dimensionality in Vertical Scales

The results of the dimensionality studies will inform the decisions concerning the grade configurations of the IRT vertical scale. If supported by the outcome of the dimensionality study, we will construct a single vertical scale for each content area, ranging from grades 3 to 11, scaled using unidimensional IRT models. It is also possible that a single vertical scale can span only certain grades and that more than one vertical scale is needed for each content area assessment.

For example, one vertical scale might articulate grades 3 to 7 mathematics and another vertical scale might articulate grades 8 and EOC high school mathematics courses. This may be particularly true as students can follow very different course sequences for high-school mathematics. In the following proposed analyses, we assume each test in a grade has a unidimensional structure. However, a vertical scale may also be possible if the test in each grade is multidimensional but with essentially the same composite of factors across grade levels (Reckase, Ackerman, & Carlson, 1988).

In any event, note that although a well-established vertical scale can be useful in assessing growth from one grade to the next, great care is needed in drawing conclusions about growth using scores from widely separated grades, especially for high stakes decisions.

Guidelines for Evaluating the Vertical Scale

Once the vertical scales have been constructed, evaluating the scales is important to establish that the tests for each grade have been properly aligned. A fundamental aspect of evaluating a vertical scale is to investigate how student scores change over grade levels by assessing changes in means and standard deviations of scale scores as well as changes in medians and selected percentile ranks.

As part of this investigation, we will examine the following statistics:

- Means and standard deviation for all grades
- Medians as well as 10th, 25th, 75th, and 90th percentiles for all grades
- Frequency relative frequency, relative cumulative frequency distributions by grade level
- Number and percentage of examinees obtaining lowest and highest scores at each grade level
- Growth measures by school types and examinee demographic characteristics

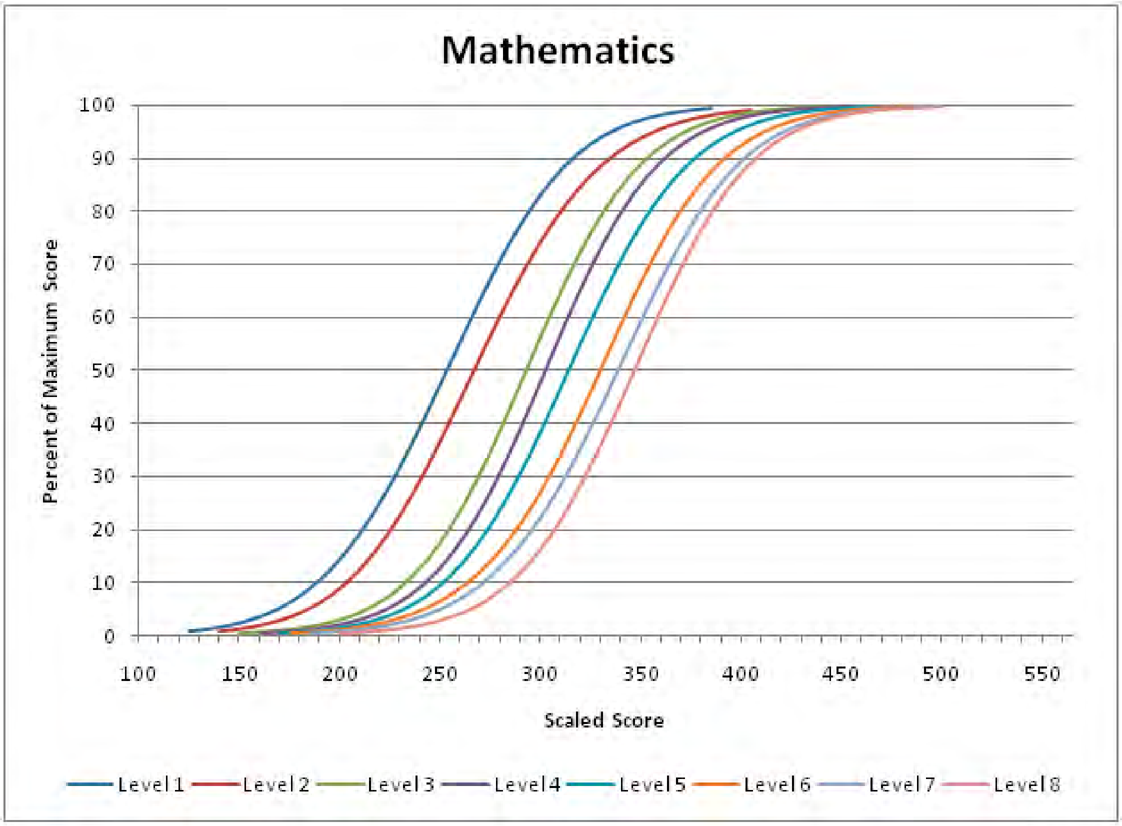
Additionally, given the goal of using vertically scaled test score gains to provide measures of growth for instructional and possibly accountability uses, we will evaluate scales with respect to departure from the ideal of equal interval properties.

As described by Briggs and Domingue (2013), a relatively straightforward approach entails examination of the gains needed to maintain the same percentile score across grades at key points of interest along the performance continuum. Specifically, compare “for each pair of adjacent grades and each scale, the ratio of the gains needed to maintain a position at the 25th, 50th, 75th, or 90th percentiles relative to the gain needed to maintain a position at the 10th percentile” (p. 14). In this framework, percentile gain ratios that deviate from 1 indicate departure from equal interval properties.

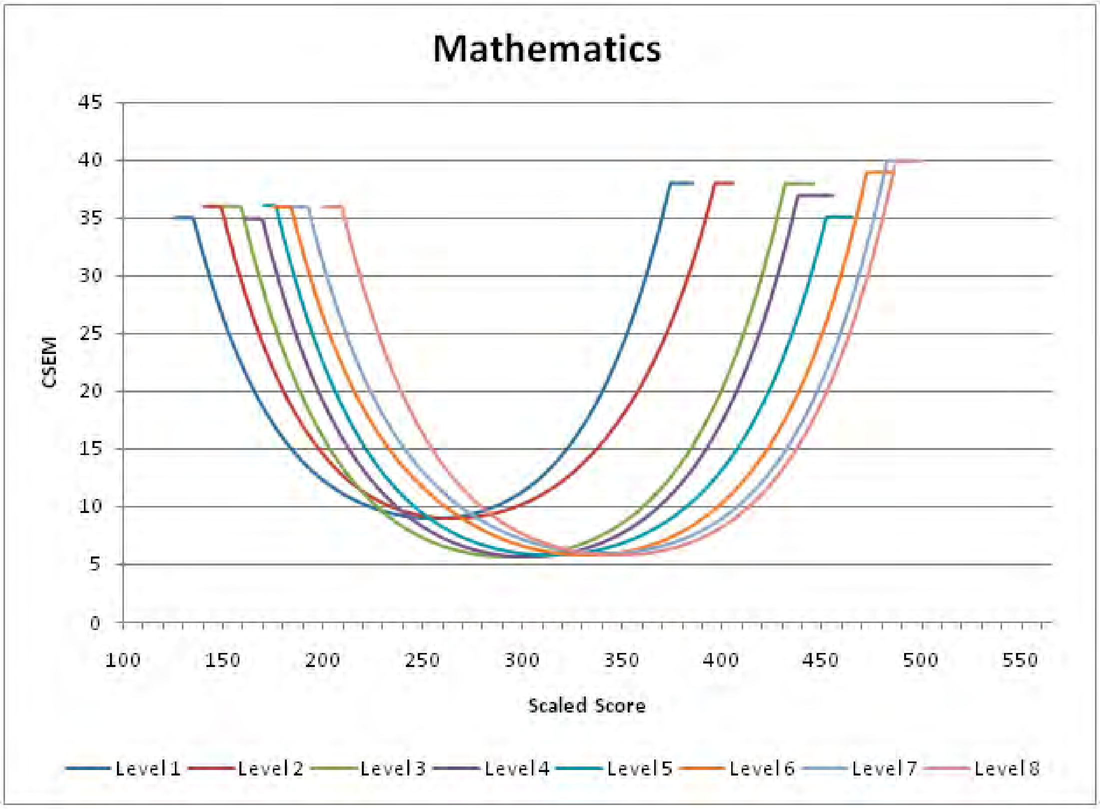
Another important component of evaluating the reasonableness of vertical scaling results is to examine the test characteristic curves (TCCs) across grades to assess whether these curves match expectations given the particular content area, the way in which the tests were constructed, and how students typically learn a particular content area across the span of grades included in the vertical scale.

We will plot and examine the TCCs for forms on the new vertical scale, for ELA/literacy and for mathematics, for overall progression in difficulty. It is expected that, in general, forms will increase in difficulty as grade level increases. This progression in the order of the TCCs on the vertical scale should coincide with the changes in test difficulty over grade level.

Likewise, we will also plot and examine the conditional standard errors of measurement (CSEMs) for tests on a vertical scale for overall ordinality. It is expected that as test level increases, the ability level at which the test is measuring most accurately will increase as well. The following figures show examples of the expected ordinality of TCCs and CSEMs for a properly functioning vertical scale. In this particular hypothetical example the tests at levels 1 and 2 have fewer items than the tests at levels 3 to 8, resulting in higher CSEMs.



Sample Mathematics Plots of TCCs. This is how we would expect TCCs plots to look for a properly functioning vertical scale.



Sample Mathematics Plots of CSEMs. This is how we would expect CSEMs plots to look for a properly functioning vertical scale.

Monitoring of the Vertical Scale

After psychometricians and content specialists have agreed on the integrity of the new vertical scales and that the tests for each grade have been aligned as expected, item parameter estimates for the operational administration will be on the common vertical scale and we will use them to construct the operational tests in the first administration.

In future operational administrations, PARCC may evaluate the vertical scale periodically using post-equated item parameter estimates to determine whether the scale is stable. This requires linking items across grades and test administrations. We recommend this be done as part of scale maintenance every five to 10 years.

We have experience in developing vertical scales and monitoring their use for a number of large-scale assessments. Once a scale has been established, we work with our clients to evaluate scale performance over time to help confirm that reported scores are following expected patterns according to the established scales and construct being measured. The same methodology is recommended for PARRC.

Special Consideration: High School Mathematics Assessments

The Common Core State Standards for Mathematics (CCSSM) are organized by grade level for Grades 3-8 and by conceptual category for High School. To assist states in aligning instruction to the CCSSM, model course pathways were developed for High School mathematics with standards organized into two sequences of coursework designed to lead to college and career readiness and to prepare students for study in more advanced mathematics courses.

Traditional Pathway

One pathway is based on organization of high school course work typically observed in the United States. This Traditional pathway includes two algebra courses and a geometry course, with some data analysis, probability, and statistics included in each course.

Integrated Pathway

The second pathway is based on a more integrated approach to secondary mathematics that is less common in the United States, but typically observed internationally. This is the Integrated pathway and includes a sequence of three courses, each of which includes algebra, geometry, data analysis, probability, and statistics.

Each pathway includes all of the standards detailed in the CCSSM, with the pathways differing only in the timing and grouping of the standards within each sequence.

PARCC will offer end-of-course operational assessments for each pathway, as articulated in the PARCC Model Content Frameworks, specifically algebra I, geometry, and algebra II for students who follow the Traditional course sequence, and mathematics 1, 2, and 3 for those taking integrated mathematics courses.

A challenge for measuring student progress toward college and career readiness in mathematics arises when students at the same point in their secondary education are taught mathematics courses that differ substantially in terms of the focus and sequence of instruction.

To address this challenge, PARCC would like to determine the extent to which items administered to students in the Traditional and the Integrated course sequences perform similarly and can be placed on the same scale to support comparisons of student achievement in the domains assessed in the respective course sequences.

As part of the Field Test Assessment administration, PARCC will investigate the degree of comparability of psychometric properties of items administered in both the Traditional and the Integrated course sequences. The purpose of this study using field test data is to determine whether items administered to students in the two course sequences can be placed on a single scale or metric. The ability to place all items on a single scale would greatly facilitate the comparison of students' degree of mastery of the domains being assessed in each course sequence.

Comparability of High School Mathematics Assessments

During the first operational administration of the High School mathematics assessments, PARCC would like to determine the comparability of the psychometric properties of items administered in both course sequences. PARCC would also like to evaluate whether the degree of comparability observed during the field test phase is maintained during the operational assessment.

In the following section we propose a replication of the research study that is planned for the field test, to be conducted during the first operational administration of the High School mathematics EOC assessments. The purpose of the study is to evaluate the comparability of the two mathematics course sequences. Sets of vertical linking item will be shared between courses in each sequence (e.g., Mathematics I and Algebra I assessments share common items, Mathematics II and Geometry assessments share common items, Mathematics III and Algebra II assessments share common items).

We will use IRT analyses to evaluate common item parameter estimates and item mapping to assess the consistency of the meaning of results for the two course sequences. Specifically, the item mapping procedure will facilitate examination of the correspondence between the knowledge, skills, and abilities (KSAs) underlying the scores in each course sequence at key points on the score scale.

The cross-sequence linking would be as follows:

- Algebra I ↔ Mathematics I
- Geometry ↔ Mathematics II
- Algebra II ↔ Mathematics III

We will conduct separate calibrations of the items comprising each EOC test form, and place the item parameter estimates and corresponding ability estimates for each pair of EOCs on the same scale using the Stocking and Lord common item linking procedure. The resulting linkage will permit cross-sequence examination of the consistency of the item difficulties.

In addition, examining the linkages at the test level will provide additional information related to the degree of comparability (e.g., equated, linked, or concorded) that can be achieved between the assessments of the two course sequences.

We will create item maps for each course that include the course-specific items and the common items, identifying each item type. The common items will provide the vehicle for aligning the items from the two courses. Criteria for location of items on the map may be based on item difficulty, a specified response probability, or a point of maximum information. For ease of interpretation, item locations and scale scores for both tests may be expressed in a preliminary reporting scale metric, through linear transformation.

For example, linear transformation using a multiplicative constant of 100 and an additive constant of 400, would convert the mean and standard deviation of thetas to 400 and 100, respectively. Key points on the scale will be identified on the map; they may be based on the score distribution, for example, the mean and one deviation above and below, or on particular scores of interest, such as performance level cut scores.

Sample item maps are presented in the following figure for two hypothetical EOCs corresponding to the first of three courses in each sequence, with the item map for Algebra I on the left and the item map for Integrated Mathematics I on the right. A reporting scale metric of mean 400 and standard deviation 100 is used. Items are labeled by item type and position within the test form. Items unique to each form are labeled with ‘A’ and ‘M’ for Algebra I and Mathematics I, respectively.

The label for items common to both tests (linking items) begins with ‘L’. For ease of interpretation, the linking items are represented as items 6-10 on each of the test forms, and items 1-5 and 11-15 are the unique items. The key scale score points are 300, 400, and 500, corresponding to one standard deviation below the mean, the mean, and one standard deviation above the mean.

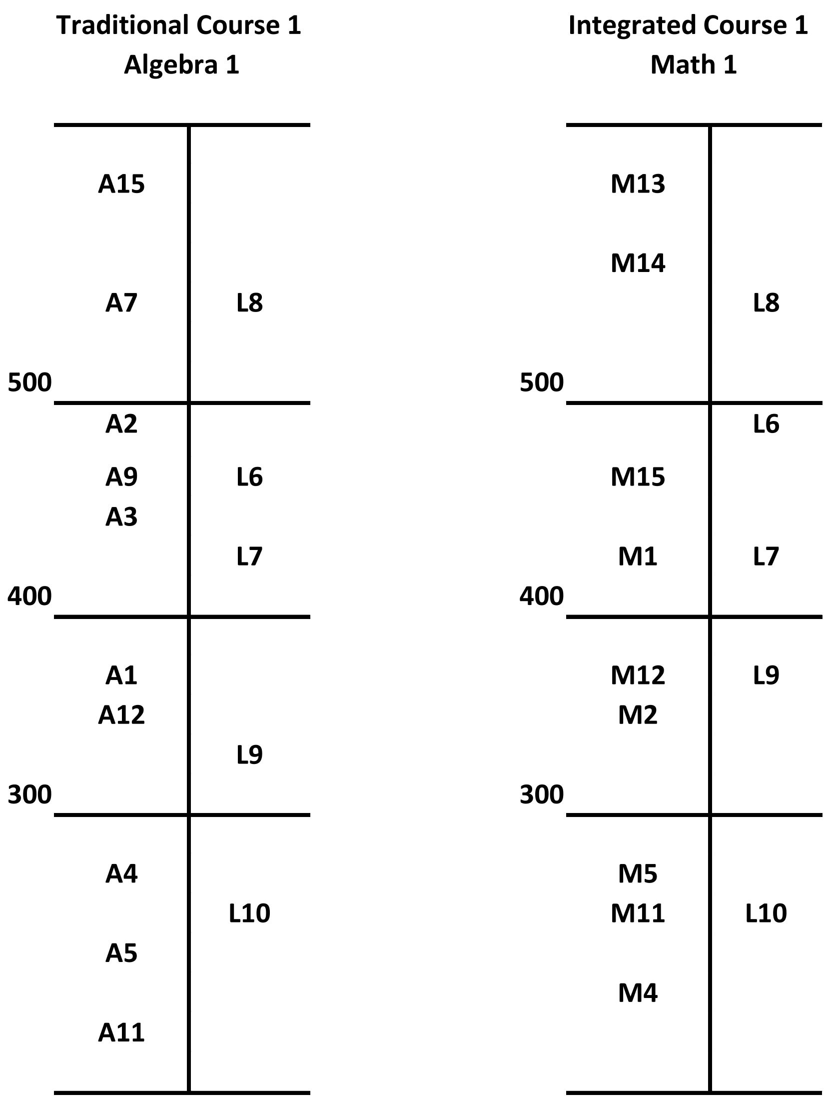
We will have content experts compare the distribution of items on each item map, and interpret the meaning of scores at key points on the scale in terms of the KSAs represented by the distribution of items in the vicinity of the score. To assist in this review, we will provide a copy of each item, information on the standards assessed by the items, the test blueprints, and other materials to support interpretation such as performance level descriptors.

The content experts will review content in different parts of the scale and interpret performance on the two tests. In particular, they will provide feedback on the comparability of the meaning of scores on the two tests, by answering questions such as “Does obtaining a 400 for Test I match what it means to obtain a 400 on Test II? What about a score of 500?”

PARCC will set performance standards following the first operational administration. To support this effort, the establishment of preliminary performance levels is expected to inform whether each performance level means essentially the same thing in terms of degree of content mastery. If the evaluation indicates that the meanings of key scores of interest for each EOC are the same, then there is support for use of the same scale.

If not, then use of different scales is recommended. Examination of each pair of courses in the two sequences may indicate that different scales are needed for the first two courses in each curriculum, but after three courses there is sufficient alignment to allow use of a single scale.

Hypothetical Item Maps for First Courses in Traditional and Integrated Sequences



This study, designed to investigate the comparability of courses between the two mathematics sequences, will result in one of several possible outcomes, arranged from most robust to least robust. Possible outcomes include the following:

1. **Pairs of Tests Can Be Equated.** The same IRT scale can be established across tests and the same item parameter estimates are obtained for pairs of common items. This finding will be supported if:
   a. Most or all of the common items perform identically between the two sequences.
   b. Both tests are unidimensional.
   c. Item mapping demonstrates that common items are closely aligned between each pair of courses.
2. **Second Order Equity Between Pairs of Tests.** This means that only some of the common items could be used as linking items, the IRT scales are different between the two sequences, however the same score scale can be established. This finding will be supported if:
   a. The two pairs of tests have comparable conditional standard error of measurement (CSEM).
   b. Item mapping demonstrates that common items are adequately aligned between each pair of courses.

      iii. **Concordance Between Pairs of Tests.** This means that the tests represent different dimensions and cannot be equated. We can construct concordance tables to 'align' scores between the pairs of tests if correlations on common items are relatively high (e.g., >0.80).

An advantage of this study is that item mapping enables in-depth examination of item performance relative to key score points of interest. A further advantage is that this study will utilize the data collected during the first operational administration of the end of course (EOC) tests. Therefore, no costs will be incurred for additional form construction or student sampling.

The disadvantage is that cross-sequence common items may be limited as well as the number of items within content, and may therefore not provide a broad sampling of the construct of measurement interest across the score range. Also, this study design does not directly examine differences in students' performance on the two assessments.

Vertical Linking of Mathematics

PARCC also requests a solution for establishing a vertical scale for grades 3–8, as well as the various mathematics EOC sequences, using operational data. This plan will investigate whether vertical scaling is feasible for PARCC assessments and, if so, which grades/courses and subjects or learning progressions are viable. The plan will also determine guidelines for evaluation of the vertical scale.

Section V.C.1, subsections D.1 and D.6 provide detailed descriptions of the plan to develop and evaluate a vertical scale for PARCC using operational data. This section focuses specifically on considerations for establishing a vertical scale that includes EOC mathematics assessments.

For mathematics, we will employ a multi-faceted approach to evaluating a vertical scale that spans grades 3–8 and EOC. We will first work with the grade-level tests for grades 3–8. Analysis of the articulation of content across EOC courses in each sequence will guide the next steps, as will the results of the mathematics EOC comparability study, described earlier in this section. These results will inform the feasibility of vertical scales for the EOCs.

Theoretically, the EOC vertical linking design would be as follows for EOC mathematics, for each sequence:

- Grade 8 ← Algebra I ← Geometry ← Algebra II
- Grade 8 ← Mathematics I ← Mathematics II ← Mathematics II

Evaluation of the vertical scale which includes EOC mathematics would follow the guidelines described in subsections D.1 and D.6. Should results show that content articulates well in the courses for the Integrated mathematics sequence but not for the Traditional courses, it may be possible to build a vertical scale using the Integrated EOCs, and then place the Traditional EOCs onto the vertical scale though horizontal linking to the corresponding on-scale Integrated EOCs. These horizontal links will have already been established, as described earlier in this section, to evaluate EOC comparability between the two sequences.

We look forward to working with PARCC to finalize plans to evaluate the comparability of mathematics end-of-course tests as well as plans for the vertical linking design.

Equating Across Years

As previously described, in year 1 we will use census data to establish the operational scale for the PARCC assessments. This will serve as the base scale to support score comparability and the reliable classification of students into one of five performance levels within and across administrations.

In the text below, we outline the three equating options requested in the RFP to maintain score comparability across years

1. Pre-equating
2. Post-equating with a calibration sample
3. Post-equating with census data

After the base scale is established in year 1, PARCC has indicated interest in options for post-equating and pre-equating across subsequent years. Post-equating would take place under the common item non-equivalent groups design. Under this design, we administer one set of items (e.g., Form A) to one group of students, and another set of items (e.g., Form B) to a second group of students.

In addition, a third set of items, common to both forms, is administered to both groups. We build this block of items, usually known as an anchor set, to represent the total test in terms of both content and statistical properties. We use the common items to identify differences in the ability of the two groups that are taking the test forms being equated.

Identification of differences in group ability enables an evaluation of differences in difficulty between forms. We examine the common items prior to equating to establish that these items are functioning similarly on both test forms. If necessary, after consultation with content experts and appropriate technical advisors, we exclude items not functioning in a similar manner for construct-irrelevant reasons on both forms from the common equating block prior to equating.

The common-item design is amenable to either classical or IRT-based equating methods. The former includes chained equipercentile equating or the linear methods of Tucker and Levine (Kolen & Brennan, 2004). We discuss classical approaches in the “Equating Tests that Do Not Have an IRT Scale” section below.

We can also apply IRT equating methods, either by calibrating all test forms to be equated jointly (or concurrently) or calibrating each separately and then using the anchor items to define the scale links that join the separate calibrations together. We propose the Stocking and Lord (1983) test characteristic curve (TCC) linking procedure to estimate these scale links that will shift the “new” form item parameter estimates onto the established, or “base” scale.

Under the pre-equated design, we assemble new forms from a bank of items that have already been calibrated and linked to the base scale. The banked IRT parameter estimates associated with the items comprising each new form are then used to produce the corresponding scale scores. We establish a calibrated and scaled bank by field-testing each newly developed item in a way that allows it to be both calibrated and linked to the base ability scale.

Items can be field-tested either by administering them alongside previously scaled operational items (embedding) or administering them to selected student groups outside of operational testing (for example the ELA/literacy PBA field test). In the latter case, the standalone field test would need to include a sufficient number of previously scaled items to serve as anchors to the bank scale.

In the following section we describe the advantages and disadvantages of each option. We also propose a series of analyses to evaluate these options upon receiving census data in year 1. We will work with PARCC to develop a detailed plan to implement the preferred equating option to provide reliable results in a timely manner for Years 2-4.

Post-Equating vs. Pre-Equating

Post-equating and pre-equating are both widely accepted equating methods that provide scale scores and resulting proficiency classifications which can be accurately compared from year to year.

Post-equating uses current item data (i.e., data obtained from the most recent administration of a test form), to produce the scale scores for that test form. To conduct post-equating, it is necessary to first accumulate sufficient, representative data from the current administration. The obvious disadvantage is that the reporting of scores is necessarily delayed until such time as sufficient data can be collected and analyzed. This may entail a delay of anywhere from three weeks to eight weeks, depending on administration volume.

A common approach for post-equating is to base the equating on a sample of students who have tested early in the administration window. Once the necessary calibrations and equating have been completed using data from this early sample of students, scoring of subsequent students can take place immediately, just as would be the case with a pre-equated design. Post-equating procedures are well documented and have been proven effective in a wide variety of K-12 testing programs.

Pre-equating is performed prior to the operational administration of the test form, using data from previous operational administrations or outside data collection events. Because each new test form is comprised of items that are already calibrated using IRT relative to the bank (or base) scale, ability estimates or expected true scores relative to this scale are immediately available.

Scoring and reporting can therefore take place very quickly regardless of when a student tests with the administration window. However, pre-equating has some disadvantages. Most notably, use of item parameter estimates established from previous test administrations may introduce instability in scores from year to year. Pre-equating assumes that all items perform in the current administration just as they did in the previous administration when calibrated to the bank (or base) scale. Post-equating assumes this is true only for those items designated as anchors, a much weaker requirement. Furthermore, the post-equated design allows this assumption to be checked and confirmed prior to equating.

To produce results that are sufficiently accurate for high stakes decisions, it is therefore critical that a pre-equated design be structured to control for potential factors that may change item performance. These include factors related to form design (item context and item position effects, form length, section breaks, etc.), conditions of administration (including speededness, administration mode, manipulatives, etc.), curriculum changes, and the simple passage of time. It is also important that the test data on which calibrations are based be obtained from large, representative samples of students, ideally under operational conditions without modification of any type to the items.

We acknowledge that PARCC prefers to use a pre-equated design in year 2 and beyond to be able to deliver data for accountability purposes under timelines needed by PARCC member states. However, PARCC recognizes that pre-equating may not provide reliable results in the early years of the assessment program and that post-equating may be necessary. Depending on the impact on the reporting timeline, there are three equating alternatives:

1. Pre-equate in year 2 and beyond (timely reporting of scores)
2. Post–equate in year 2 and beyond using a calibration sample (relatively timely reporting of scores depending on calibration sample)
3. Post–equate in year 2 and beyond using a census data (least timely reporting of scores)

Timelines for the pre-equating scenario, as well as for the post-equating using a calibration sample would allow for delivery of scale scores to PARCC states by the end of their respective school years. In the third scenario, where post-equating would be based on census data, either of the following two options would allow for delivery of scale scores to PARCC states by the end of their respective school years:

1. Provisional scale scores could be reported, based on post-equating on calibration data, with the caveat that the scores may be adjusted within an allotted timeframe—dependent on post-equating with census data
2. A policy decision is made to define “census data” as some percentage less than 100 percent of the data, perhaps 90 percent, that would enable post-equating results to be reported in the desired timeframe

Our aim is to produce accurate and timely scoring, and we will collaborate with PARCC and its TAC to identify the most viable option to support the goals of the assessment system and the needs of the PARCC member states.

Evaluating Equating Options

To evaluate equating options for the year 2 operational administration, we will use data from the year 1 administration to investigate the stability of item statistics and resulting scores relative to the previous administration (the spring 2014 field test). Although not directly comparable because of differences in the conditions of administration and associated stakes, the results of the analyses should provide a proof of concept to inform the choice of equating options in year 2.

Specifically, we will investigate the three equating options (pre-equate in year 2 and beyond, post-equate in year 2 and beyond using a calibration sample, post-equate in year 2 and beyond using census data) by addressing two research questions. These questions and the analyses needed to answer each are described below.

Question 1) Are the item parameter estimates produced at different administrations sufficiently stable to support pre-equating? This can be gauged by observing the magnitude of differences between item parameter estimates obtained from the field tests and those obtained from the year 1 census data.

Required Analyses

The operational forms administered in year 1 are constructed based on the field test statistics. Item and test-level statistics obtained from the year 1 census data will be therefore compared to the field test statistics. For the purpose of this comparison, the post-equated alternative for year 2 will be simulated by equating the year 1 item calibrations to the field test bank scale. We will examine the following statistics:

- Item-total score biserial correlations
- IRT parameter estimates and item characteristics curves (ICCs)
- Test characteristic curves (TCCs)
- Conversion tables based on field test statistics (“pre-equated”) and those based on “post-equated” item statistics

We can evaluate the stability of item-total correlations with scatter plots and 95 percent prediction bands. If a large number of items fall outside of the prediction bands, we expect the equating results to differ significantly.

We can evaluate the stability of IRT parameter estimates and associated ICCs with methods commonly used to evaluate anchor item parameter drift in equating. With the Rasch model, an item parameter difference of 0.3 logits or more is considered to be a large difference.

Using the ICC method, the root mean square difference (RMSD) between the two ICCs of the same item can be calculated; RMSD values greater than 1.0 are generally considered to indicate significant differences between item parameter estimates.

We will also examine the differences between score conversion tables that are produced under pre- vs. post-equating methods, and the resulting impact on students' performance level classifications. We will look at the largest difference and the average differences in scale score conversions, as well as how many students are affected by these differences.

At some points on the score scale, a difference of one scale-score point may not affect students. However, at points near performance cut scores, a small difference in the raw-to-scale score conversion could result in a meaningful difference for individual students and in the percentages of students assigned to performance categories, particularly if many students have scores near that cut score. We will conduct a holistic review of differences between conversion tables to determine the significance of these differences and the impact to students.

A pre-equated design may not be appropriate during the first several years of operational administration, particularly if a large number of items perform significantly differently compared to the base year administration. To allow time for a new testing program to stabilize, a post-equated design is often preferable. Post-equating following each operational test administration in the early years will produce reliable equating results.

Question 2) Is the early return sample sufficiently representative with regard to relevant characteristics to support post-administration calibration and equating? We can judge this by comparing year 1 calibration results from designated early samples with results from the census data.

Required Analyses

We will compare calibrations based on early return samples to calibrations from census data. To reflect best practice for future operational administrations, we will identify early return samples that reflect the total test population. The target size for the early return samples will be determined in consultation with PARCC and its Technical Advisory Committee. We will compare calibrations from these samples with those from the census data for each subject and grade/EOC test.

Specifically, we will evaluate differences in the item parameter estimates and the conversion tables, obtained from the calibration samples and the census data using all of the analyses described for question 1, above. We will also investigate differences in classification of students into performance categories, based on the early-sample and census-sample equating. In this study, we will consider post-equating with census data to be the "true" equating function or "gold standard." These comparisons, along with the impact on the reporting timeline, will inform the decision of whether early return samples can be used for post-equating to provide reliable results.

If post-equating with early return samples is determined as the equating method of choice in year 2 and moving forward until pre-equating becomes feasible, we recommend use of a targeted early return sampling plan to provide for stable, representative samples. Prior to start of analyses, we will analyze the sample characteristics (demographics, and prior year performance— school or student if available) to evaluate the year-to-year consistency in representation of the testing population. If the characteristics of the early return samples fluctuate significantly from year to year, we can use a post-stratification method to construct consistent samples for equating purposes and to facilitate evaluation of the equating results.

Preparing the Transition to Pre-equating

PARCC would like to eventually transition to a pre-equated design; therefore, in the early years when post-equating is used, we must give special considerations when building and administering the forms to verify that the item statistics obtained are reliable for pre-equating.

Within the pre-equated design, items are necessarily assumed to perform identically during operational administration as when these items were first tested and calibrated. To this end, under the pre-equated design, if item parameter estimates are obtained when items are field tested, it is important that field test items be administered under the same conditions and at the same time in the school calendar as when the operational tests are administered.

The most efficient approach is to embed some number of new, un-scaled items within the operational tests. These new items are then calibrated and linked to a reference or “bank” scale. The anchor for this linking can be either the core set of operational items or an embedded set of previously-scaled items that is external to the operational core.

Equating Tests that Do Not Have an IRT Scale

We can implement the post-equated design without IRT, should we find that IRT models work poorly with some test components, such as PBA writing tasks. We can use the same common-item non-equivalent groups design as described earlier with classical equating methods (e.g., chained equipercentile or Tucker).

ETS has developed many of these methods that are used throughout the industry today for test equating (Angoff, 1971; Holland & Rubin, 1982), and we strive to continue leading equating research (Puhan, Moses, Grant, & McHale, 2009).

Another alternative is to employ a hybrid approach, where some test components are scaled and equated using IRT while classical equating methods are applied to other components. This type of hybrid will be especially relevant for equating ELA/literacy if it is determined that IRT can be used to scale the reading component but not the writing component. In this case, we will equate the reading and writing components separately, and we will then equate the ELA/literacy composite score, by either common item or equivalent group approaches.

Evaluation of Linking Items

When new test forms are built or new items are field tested, a set of items may be selected to serve as linking items for equating across forms and administrations. The following sections describe analyses and corresponding guidelines to evaluate the integrity of linking items, a proposed solution for monitoring rater drift in constructed-response linking items, and a sampling plan for hand scoring constructed-response items that serve as external links.

Evaluation of Linking Items Using Classical Item Analyses

Classical item analyses involve computing a set of statistics, for every linking and non-linking item, in each form. Each statistic is designed to provide some key information about the quality of each item. The criteria for evaluating the following classical item statistics are provided in Section V.C.1.B.

- **Classical Item Difficulty Indices (or *p*-value, selected-response, and constructed-response items).** For dichotomously scored items, this statistic indicates the proportion of students who answered the item correctly. For polytomously scored items, this statistic represents the average item score or the portion of the maximum obtainable score.
- **Percentage of Students Choosing Each Response Option (selected-response items).** These statistics indicate the percentage of students who select each of the answer options and the percentage that omitted the item. We will flag items for review if a greater number of high-performance students choose a distractor rather than the keyed response.
- **Item-Total Biserial Correlation (selected-response and constructed-response items).** This statistic describes the relationship between performance on the specific item and performance on the total form (PBA and EOY). For constructed-response items, the item-total biserial correlation is used; for selected-response items, the item-total correlation is the polyserial correlation.
- **Distractor-total correlation (selected-response items).** This statistic describes the relationship between selecting an incorrect response for a specific item and performance on the total form (PBA and EOY). The values of this correlation will be compared and contrasted with the discrimination index (see 4.1.3: item-total correlation above).

Evaluation of Linking Items Using IRT

The linking process is iterative and involves an inspection of differences between the transformed item parameter estimates (based on year 2 of the operational administration) and the reference item parameter estimates (based on year 1 of the operational administration) for each of the linking items.

We will flag items that show large differences between transformed and reference item parameter estimates for review by psychometricians and assessment development specialists. PARCC will review any items recommended for removal from a linking set. If approved, we will eliminate these items from the linking set and we will re-run the Stocking and Lord scaling (see Section D.1) to re-estimate the transformed item parameters.

We will further evaluate the linking items using the similar procedures as described for the evaluation of vertical items. These criteria have produced reasonable results over time, and have been used satisfactorily for other testing programs.

In addition to evaluating individual linking item performance, we will examine correlations between the transformed and reference item parameter. In general, high correlations are expected. These correlations tend to be slightly higher for mathematics than for ELA\Literacy, possibly because ELA/literacy items are passage-dependent and more susceptible to context and position effects. .

Constructed-Response Linking Items

ETS psychometric staff will work with the Pearson Performance Scoring Center to collect data that can be used to evaluate the consistency of scoring across years. For ELA/literacy, current plans are to link the PBA sections from year to year through the EOY sections. For mathematics, constructed-response items are likely to be part of year-to-year linking sets. However, for both ELA/literacy and mathematics, it is possible that constructed-response items will be in operational test forms used across years. Thus, a mechanism for evaluating year-to-year rater drift should be developed not only for constructed-response linking items, but for all constructed-response items used across administrations.

In order to monitor the consistent scoring of constructed-response items used from year to year, Pearson will save and reuse reader training materials (anchor, practice, and qualifying sets), validity, and calibration sets. Handscoring leads will be required to reuse the same validity and calibration responses from the previous administration, although validity and calibration may be presented in a different order based on scoring trends. New calibration or validity may be added only with the approval of the Partnership Manager. In this way scorers will receive the same training from year to year, and will be evaluated according to the same standards.

Pearson will work with ETS to facilitate the documentation of trend data regarding the validity of scoring based on introducing the same calibrated responses (i.e., the same validity papers) from year to year. ETS will compile the year to year validity data and include these data in the annual technical manual.

Claim and Sub-Claim Level Scores: Raw Scores and Domain Scores

The master claim that students are “On Track” or college and career ready reflects the overall goal of the Common Core State Standards and Model Content Frameworks—to prepare students for college and careers, and specifically to verify that students have the skills and understandings required for success in higher education and in their careers. The student’s progress towards this essential goal will be reflected by the student’s overall performance on the summative components (both the Performance-Based Assessment and End-of-Year Assessment) of the PARCC Assessment System.

Scores for sub-claims will provide additional data in support of the Major Claims and will help educators identify the skills in which the students are particularly strong and those in which they are particularly weak. This information will enable them to target their instruction where it is most needed.

Estimating Sub-Claim Level Scores

In general, there is often tension between trying to maximize the amount of detailed information that can be drawn from a test (say, by increasing the number of subscores) and having a sufficient number of items that contribute to each score so as to provide useful information. Useful subscores support appropriate actions by having appropriate statistical accuracy and reliability for their intended use, as well as sufficient numbers and types of observations so results are generalizable. For PARCC sub-claims, we expect to obtain such information based on the field test analyses. Possible uses of sub-claim scores include:

- Feedback for students and parents about areas of strength or weakness
- Feedback for schools to inform instruction and school improvement plans
- Information to inform high stakes decisions (e.g., admission or placement)

In general, the higher the stakes of decisions based on sub-claim scores, the greater the need to verify a sufficient number of items and appropriate items per sub-claim score and the more care needed when communicating the meaning of the sub-claim scores.

The intended use of sub-claim scores affects the appropriate metric for reporting these scores. Sub-claim scores used for low-stakes decisions may be reported as a percent correct which is a very straightforward metric to understand; the disadvantage is that these scores are not comparable over test forms or grades. Sub-claim scores can also be reported using an IRT scale score metric that enables comparison of scores across test forms.

A related IRT approach is to use expected true scores (based on estimated theta and item parameters). Expected true scores can be estimated for the entire pool of items for a given domain even if each student takes only a portion of those items. The RFP had indicated that the Partnership is considering this method for reporting at the claim and sub-claim level.

Although this approach facilitates comparisons of scores over test forms, it can mask true dimensional differences among content areas that make up different claim or sub-claim areas.

Reporting Metrics for Sub-Claim Level Scores

Below we discuss five possible options that might be used to report sub-claim level scores:

1. A percentage of the maximum possible score on the tasks presented to the student
2. A percentage of the maximum possible score on some specified set of tasks that is the same for all students
3. The student's percentile rank in a specified norm group
4. A scale based on the mean and standard deviation of the scores of a specified norm group
5. The same vertical scale used for the total scores

Option 1 is simple and direct, but the scores would not be comparable across forms of the test. The items contributing to each sub-claim could vary substantially in difficulty from one form to another. As a result, comparisons of the same sub-claim on different forms of the test could be misleading.

Option 2 has the advantages of Option 1 and avoids the problem of comparing percentages that refer to different sets of items. In this option, the items could change from year to year, but the specified set of items that serves as the basis for score reporting (the "base test") would remain constant. This is a variation on the idea of an expected true score on an entire pool of items comprising a domain.

In this case, the reference domain is the set of items on the base test. Scores would then be comparable across years as well as across forms within a year.

Option 3 produces a score that is easy to interpret. It has the advantage and the disadvantage of describing the student's performance by comparing it with the performance of other students.
Percentile ranks have the additional disadvantage of being overly sensitive to small differences near the middle of the scale, where the score distribution is dense; a small difference in performance can result in a large difference in the students percentile rank.

Option 4 is also a score that requires normative interpretations. It avoids the problem of being overly sensitive to small differences in performance. However, the scores that result are not as easy for parents and teachers to interpret. The scale will appear to be (and for many practical purposes, will be) essentially arbitrary.

Option 5 has the advantage of producing sub-claim scores that are linked to the total scores. This advantage can also be a disadvantage, because the linking process will remove any real differences between sub-claims in the performance of the group of students as a whole. Areas of strength and weakness would be defined in relative terms. A student who is only slightly weak in a skill area where most students are very weak would appear to be strong in that area.

ETS is willing to pursue any of these potential methods that PARCC specifies, but our preference is Option 2: reporting the sub-claim as a percentage of the maximum possible score on a “base test” for that sub-claim. PARCC could make the specific tasks on the base test available to the public, to clarify the meaning of this percentage.

Number of Items and Score Points for Adequate Reliability

The reliability of any score depends on the number of independent observations contributing to that score. For a test made up of dichotomous items, that number is the number on the test. For a test consisting of items on which partial credit is possible, the relationship between score points and reliability is not so simple. A task that generates a score on a scale of 0 to 5 will yield more reliable information than a single dichotomous item, but less than the sum of five dichotomous items. Changing the score scale on a task from five points to 10 points will produce, at best, a small increase in the reliability of the task score. Adding a second five-point task will produce a much larger increase in realiability. .

There are other complicating factors in this discussion. Tasks with partial-credit scoring are generally more time-consuming than dichotomous items. They are also less likely to be affected by a student’s luck in guessing at the answer.

While most testing experts would be reluctant to specify a “satisfactory level of reliability,” most would agree that the level of reliability that is satisfactory depends on how the scores are to be used. We present our recommendations for computing subscores that are based on the following assumptions about the use of the subscore:

- No high-stakes decisions about individual students will be based on the subscores.
- The subscores of individual students may be used by teachers for diagnostic purposes.
- The subscores of groups of students (e.g., those in a school or a school district) may be used for making decisions about curriculum and instruction.

For subscores computed entirely from multiple-choice or other dichotomous items, ETS recommends that each subscore reported for individual students be based on at least 12 items.

For subscores computed from tasks scored on a three-point scale (e.g., 0, 1, 2), ETS recommends that each subscore be based on at least 6 separate tasks.

For subscores computed from tasks scored on a four-point scale (e.g., 0, 1, 2, 3), ETS recommends that each subscore be based on at least 4 separate tasks.

For subscores computed from tasks scored on a five-point scale (e.g., 0, 1, 2, 3, 4) or a six-point scale (0, 1, 2, 3, 4, 5), ETS recommends that each subscore be based on at least 3 separate tasks.

For subscores computed from tasks scored on a seven-point scale or any finer scale, ETS recommends that each subscore be based on at least 2 separate tasks.

For subscores comprised of a mix of item formats, we recommend that the subscore to be based on at least 16 total points.

We note that these recommendations are based on our experience and professional judgment, and we welcome discussion with PARCC and their technical advisors on the tradeoffs involved in reporting scores at the sub-claim level.

Growth Scores

As specified in the RFP, PARCC is committed to reporting two types of common measures of annual progress—absolute and normative—that describe annual changes in student performance. PARCC is also committed to placing any use of PARCC assessment data to inform accountability and evaluation decisions at the discretion of each participating PARCC state. In line with PARCC's commitments, the measures of annual progress that ETS proposes below are only discussed with regard to their descriptive or predictive (i.e., growth-to-standard) interpretations—not causal or value-added interpretations, which tend to require stronger assumptions and more considerations in model selections.

Normative Measures

As stated in the RFP, a "normative" measure of annual progress describes a students' academic progress from one year to the next in relation so his/her academic peers. Normative measures are thus useful in answering questions like: "How does my academic progress compare with the academic progress of my peers?" Currently, the most widely used normative measure by state accountability programs is Betebenner's (2009) Student Growth Percentile (SGP) model, which PARCC has expressed interest in using as its normative measure of annual progress.

ETS has extensive experience with SGPs as several ETS researchers and psychometricians have been at the forefront of investigating their properties and thus would be able to offer their vast expertise to PARCC and its Ad-hoc Committee on Growth Metrics on using the SGP model. We believe these research efforts provide ETS with incomparable institutional knowledge and expertise on the estimation and use of SGPs in accountability programs.

Betebenner's (2009) nonlinear, quantile-regression-based SGPs can easily be estimated using the "SGP" package (Betebenner, Van Iwaarden, Domingue, & Shang, 2013) in the open-source statistical package R (R Core Team, 2013). Although this approach is easy to program, it is computationally intensive and may be too nonparametric, or data-driven, in the sense that outliers may have undue influence on resultsthey can over accommodate outliers by being "attracted" to them. We recommend that PARCC consider two simpler alternative methodologies:

- (1) use linear parameterizations of the quantile regressions that would still model heteroskedasticity in the data but are less data-driven than the b-spline parameterizations
- (2) use Castellano and Ho's (2013) "percentile rank of residuals" from a single conditional mean linear regression that are very easy to implement and have been found to produce very similar results as operationally-constructed SGPs but rely on more assumptions, such as homoscedasticity

Operationally, the median is used to describe the normative performance of a group of students (e.g,.school, district, subgroup, state) following Betebenner's recommendation (2010). However, Castellano and Ho (under review) and Castellano (2011) found that the mean may be a preferable, more stable measure of central tendency as it has less sampling variability, is more invariant to monotonic transformations of the test scale, and provides more comparable group rankings with other aggregate-level conditional status measures. We therefore recommend that PARCC consider the mean as a viable aggregation function for reporting group summaries of SGPs.

As a normative measure, SGPs depend on the norming group used in their estimation. In the RFP, PARCC stated that depending on the reporting level, norming groups will be drawn from (1) the entire PARCC student population or (2) individual states. Attachment R also explains that publicly reported comparisons will include comparisons among PARCC subgroups as well as PARCC-to-State and State-to-State comparisons. For all of these publicly reported comparisons, we recommend that the norming group be drawn from the entire PARCC student population. PARCC-to-State and State-to-State comparisons of mean or median SGPs will not be comparable if the SGPs for students in each state are estimated using state-specific norming groups.

Overall, ETS recognizes the appeal of using SGPs as a normative measure of annual progress. However, we recommend that PARCC consider the percentile rank of students' absolute gains as an alternative. That is, rank order a cohort's gain scores and assign each student the corresponding percentile rank of their gain score. This metric answers the question, "How does my academic progress compare with the academic progress of all my peers?" as opposed to the question the SGP model addresses: "How does my academic progress compare with the academic progress of my peers who started at the same place as me?" Using the percentile rank of absolute gains as the normative measure would allow for a more coherent growth reporting system as it follows naturally from the use of gain scores as the absolute measure of annual progress.

This percentile rank measure relies on a vertical scale, but assuming that the investigations PARCC is undertaking ultimately support the development of useful vertical scales, it will be logical to take advantage of those scales when reporting measures of annual progress.

Absolute Measures

An “absolute” measure of annual progress allows for interpretations about how much a student has learned from one year to the next in relation to a construct that spans multiple grades. That is, absolute measures are most meaningful when a “learning progression” perspective on construct and item development is used, rather than a “domain sampling” approach (Briggs, 2013). Common absolute measures that afford such growth description interpretations are the gain score and categorical model both of which are based on gains in performance between adjacent grade levels but are expressed on different metrics (Castellano & Ho, 2013a). Gain scores are expressed as the difference between a student’s score in her current grade level and her prior grade level and thus require a vertical scale spanning all grade levels of interest. Similarly, categorical models (e.g., a transition matrix or value table) quantify the gains in performance in relation to the change in performance levels from one grade to the next, and thus require careful articulation of performance levels within and across grades.

Both of these approaches have the appeal of computational simplicity and low data demands in that they only require two consecutive years of linked student test score data. This makes missing data less likely, allowing reporting absolute measures of annual progress possible for the majority of students. However, the utility of these measures depends, to a large degree, on the extent that the metric of interest (i.e., vertical scale for the gain score model; articulation of within- and across-grade performance levels for the categorical model) can support meaningful cross-grade comparisons. In other words, determining the extent that an increase in X points or Y performance levels provides meaningful interpretations about the knowledge, skills, and abilities students have gained over time.

For the purpose of reporting an absolute measure of annual progress on student score reports and summarizing aggregate performance, ETS recommends the use of gain scores, as they allow for finer grain cross-grade comparisons and are easily aggregated to average gains for any group of interest (e.g., schools, districts, states, subgroups).

As mentioned earlier in this section, a fundamental aspect of our evaluation of the developed PARCC vertical scales is to investigate how student scores change over grade levels. We will conduct such an investigation using the following criteria:

a. Examine item difficulty and discrimination of common items between grades; items should be less difficult at higher adjacent grades.
b. Examine common item DIF results for students from subgroups (e.g., race/ethnicity, gender, SWD, and EL) to determine if an item is functioning differently in an adjacent grade.

c. Compare distributions of vertically scaled scores across grades, within same or similar school districts; looking for reasonable changes in distributions between grades.

d. Evaluate the following statistics:
   i. Means and standard deviation of scale scores for all grades;
   ii. Effect size indices based on mean scale score differences for adjacent grades;
   iii. Extent to which mean scale scores increase steadily across grades;
   iv. Medians as well as 10th, 25th, 75th, and 90th percentiles for all grades;
   v. Frequency, relative frequency, relative cumulative frequency distributions by grade;
   vi. Number and percentage of examinees obtaining lowest and highest scores at each grade.

e. Evaluate test characteristic curves (TCCs) for overall progression in test difficulty across grades.

f. Evaluate conditional standard errors of measurement (CSEMs) for overall ordinality; expected that as grade increases, the ability level at which the test is measuring most accurately will also increase.

g. Evaluate scale performance across grades (i.e., ‘d’, ‘e’, and ‘f’ in step 5) for subgroups of interest (e.g., race/ethnicity, gender, SWD, and EL).

Gain scores, in particular, rely heavily on vertical scales having equal-interval properties so that a difference of X points carries the same meaning at the bottom of the scale as in the middle and top of the scale. ETS will thus also implement Briggs and Domingue’s (2013) recommended approach for detecting departures from the ideal equal-interval scale by examining scores needed to maintain the same percentile across grades. Evaluation and ongoing maintenance of the vertical scale will afford more robust absolute growth interpretations.

Even with a strong vertical scale tied to assessments built to assess a single construct across a span of grade levels, a reported gain score of X points in and of itself may be difficult for interested stakeholders (e.g., students, parents, teachers, administrators, policymakers) to interpret. We recommend that PARCC consider providing supplemental information in student score reports and group summary reports of aggregate performance regarding interpretations of gains. For instance, student score reports could include the student’s currently attained performance level along with its descriptor and the student’s performance level from the previous year with its descriptor. Such information would help students, teachers, and parents understand what knowledge, skills, and abilities students have gained over the year.

Gain scores for a group of students (e.g., school, district, state, subgroup) can easily be aggregated by averaging to determine the group's mean change in performance from one grade to the next. Like at the student-level, gain statistics for groups may not be readily interpretable. Again, supplementing these mean gains with interpretative information about the knowledge, skills, and abilities students have at various places on the vertical scale—be it through the performance level descriptors or item mapping—would be useful.

Growth towards College and Career Readiness

The discussion of both the absolute and normative measures of annual progress in this section has been related to growth description interpretations. Given PARCC's emphasis on ensuring students are college and career ready by the end of high school, growth prediction or growth-to-standard interpretations may also be desirable to determine if students are on track to reaching this target. Both the absolute gain-score model and normative SGP model can also produce measures that afford growth-to-standard interpretations. We review the advantages/disadvantages of using each of these approaches below.

The SGP model's growth-to-standard component involves determining the minimum SGP a student needs to maintain to reach a standard, such as the College and Career Readiness (CCR) cut score, X years into the future. The needed SGP can then be compared to the student's current SGP to classify students as on track or not. Such student growth projections require having data from a cohort of students who have already reached the standard of interest. For instance, to determine the minimum SGPs that students in the current cohort of fifth graders must maintain to reach a particular cut score in grade 8, data is needed for a cohort of students who have already reached grade 8 (and have prior data for at least grade 5).

In the RFP, PARCC notes that they will report growth measures from the second year of operational administration onwards. If this includes on track growth classifications, for the first few years of operational administration, limited projections will be possible. For instance, in the second year of administration, only projections one year into the future will be possible. Moreover, these projections can only be based on one year of test score data. In general, the SGP model will only provide projections with one-year time horizons and using only current scores—no prior scores—in the second year of administration, and until, a cohort of students has gone through the PARCC assessment system from grades 3 to 8, limited time horizons and number of prior scores will be able to be used to make growth-to-standard predictions.

Another issue to consider in using the SGP model for on-track classifications is that it relies on the assumption that students will maintain their SGPs into the future. However, in practice, students' SGPs over two consecutive grade levels tend to be uncorrelated, meaning that if a student has a high SGP in grade 5 based on her grades 3 and 4 scores, her SGP in grade 6 based on her grades 3, 4, and 5 scores, could be high, low, or moderate (Castellano & Ho, in progress). Thus, a viable alternative is the projection model, which instead of assuming that students maintain their conditional status into the future, simply predicts the expected (average) score for students given their observed current and prior scores.

This model has been found to have the highest predictive accuracy (Hoffer et al., 2011; Castellano & Ho, 2013a, Castellano & Ho, in progress). Thus, if the main goal is to provide students with the best prediction of whether they are on track to college and career readiness, ETS recommends using the projection model over the SGP model's growth projection percentiles.

The absolute gain score model can be extended to a trajectory model to afford growth-to-standard interpretations. The trajectory model assumes that students maintain their current gain score (from last year to this year) into the future; that is, that the student will continue to gain the same number of score points each year. This is a linear assumption of continued annual progress. Like the SGP model, this assumption may not bear out in practice, but it has the appeal of being easily understood.

The trajectory model has the benefit over the SGP model that starting from the second year of operational administration, it could be used to make projections any number of years into the future. This follows from the fact that, unlike the SGP model, the trajectory model only requires using two consecutive years of data to predict future performance, and it does not rely on having data for another cohort of students who have already reached the target grade of interest. It is important to note, however, that the more years into the future predictions are made, the linear assumption of gains over time becomes less defensible (i.e., it might not be considered likely that students maintain the same gains from say grades 3 to 4 as from grades 7 to 8). In addition, longer time horizons rely more heavily on equal-interval properties of the vertical scale across the entire grade span.

It is valuable to point out that the use of the trajectory model to make on-track classifications along with the related gain scores as absolute measures of annual progress and percentile ranks of gain scores as the normative measure of annual progress would provide the most coherent growth reporting system. However, the choice of absolute and normative descriptive measures of annual progress as well as predictive measures involves weighing several different criteria, such as transparency, statistical rigor, and classification accuracy, to name a few.

Scope and Duration of Longitudinal Data

Estimating and reporting absolute and normative measures of annual progress requires a well-maintained longitudinal database of student test scores for students in all participating PARCC states. Unique student identifiers are fundamental to verify that students' scores can be linked across grade levels. Given that there are multiple participating PARCC states, unique PARCC student identifiers can be created by stringing together unique state identifiers with the state-provided student identifiers. Although PARCC does not support the public reporting of individual school or school district aggregate measures of annual progress, including school and school district identifiers in the longitudinal database would further aid in tracking students over time and enable us to provide results privately to each school/school district. To publicly report PARCC-to-PARCC subgroup comparisons, student subgroup

classifications (e.g., student with disability, gifted/talented, English Learner, ethnicity, gender, etc) will also need to be linked to each student.
These subgroup classifications may vary over time (i.e., a student can move from EL to non-EL status over time); thus, student data on these classifications will need to be collected each year.

Requirement

V.C.1.E. Calibrating Field Test Items

Response

To maintain a sufficiently large item bank for future operational summative assessments, as requested in the RFP, we will be field testing additional items in each summative assessment. Field test items will be evaluated using both classical and IRT analyses. The purpose of calibrating field test items is to obtain item parameter estimates to support on-going operational administrations as pertains to test form construction and the associated scores. Further, calibration of field test items is an absolute necessity for pre-equating.

Classical Item Analysis

Before calibrating any field test items, ETS will perform a series of classical item analyses of each item. The purpose of these analyses is to identify any field test items that should not be included in calibrations and subsequently used as operational items. As described earlier, classical item analyses include calculation of the following statistics: p-values, item-total correlations, distractor-total correlations, high percentage of omits, and high percentage of “not-reached.” Additionally, CR items with a low percentage of students attaining a particular score point are examined.

The analysis for each selected-response item will include the estimation of a response curve for the correct answer and each wrong answer option. The analysis for each constructed-response item will include the estimation of a response curve for each threshold (e.g., item score at least 1, at least 2, etc.). The following criteria are often used to flag items as problematic:

- Items with p-values above or below a specified threshold (e.g., above 0.95 or below 0.25)
- Item-total correlations below a specified threshold (e.g., 0.20)
- Distractor-total correlations above a specified threshold (e.g., 0.00)
- Greater number of high-performing students choose a distractor than the keyed response
- High percentage of omits (e.g., greater than five percent)
- High percentage that do not reach the item (e.g., greater than five percent)
- For constructed response items, items with a low percentage of students obtaining a score point (e.g., less than five percent)

The purpose of these analyses is to identify any field test items that should not be used as operational items, for any of the following reasons:

- The item is too easy or too difficult to provide useful information to differentiate the performance of the students for whom it is intended;
- The statistical relationship between the item and the criterion is so weak that the item will not be useful in measuring the construct of measurement interest;
- There are other problems with the item (e.g., content error, confusing presentation)

DIF Analyses

As described earlier, DIF statistics will be calculated to identify those items for which identifiable subgroups of students (e.g., males, females; white, Asian; different states, etc.) with the same underlying level of ability have different probabilities of answering correctly. Two methods will be used to flag items for further examination. The Mantel-Haenszel (MH) statistic and the logistic regression approach are complimentary in nature and can be used together to maximize the detection of DIF. Items flagged for potential DIF will be reviewed by content experts and bias/sensitivity committees to determine the source and meaning of performance differences.

IRT Analyses

The purpose of Item Response Theory (IRT) calibration and scaling of the field test items is to place the items, within each grade, on the PARCC summative assessment scale. PARCC has expressed interest in developing pre-equated assessments. The calibration of field test items is critical to achieve this goal. As described in Section V.B.2.C.1 (forms construction), field test items for the ELA EOY and mathematics EOY and PBA will be embedded in the operational test forms. Field test items will be calibrated using the common item linking method (horizontal linking). This method uses the operational items that have known item parameter estimates on the PARCC reporting scale to serve as linking items to place the field test items onto the PARCC scale.

We will use the Stocking-Lord linking method in placing the newly calibrated field test items onto the PARCC reporting scale. The ELA PBA items will be administered during the EOY administration (please see the next section).

As stated in the RFP, field test items that require handscoring will be scored with a representative sample of 1,500 PARCC students to minimize costs. Given the collective size of the student population of PARCC states, it is likely that each form (with embedded field test items) will be administered to far more than 1,500 students.

Therefore, we propose to develop a sampling plan in consultation with PARCC, its Technical Advisory Committee, and Pearson (as the performance scoring contractor) to yield a targeted sample of 1,500 valid cases from among all students who respond to each unique field test CR item. A sample of 1,500 students is typically considered sufficient for fitting a unidimensional IRT model when model assumptions are well met.

In light of PARCC's interest in possibly moving to a pre-equated design, larger sample sizes may be warranted to obtain more stable item parameter estimates in support of reliable scores. However, the costs associated with increased performance scoring would also have to be considered. We will work with the PARCC and its Technical Advisory Committee to best meet the goals and purposes of the PARCC assessments.

ELA/Literacy PBA Standalone Field Tests

The length and design of the PARCC ELA/literacy PBA component precludes the possibility of embedding field test items into operational forms. Consequently, there will be a standalone field test of the ELA/literacy PBA tasks in which field test forms comprised of single tasks administered to representative samples of students in conjunction with the administration of the EOY component at each grade level.

Three types of tasks comprise the ELA/literacy PBA: Literacy Analysis Task (LAT), Research Simulation Task (RST), and Narrative Writing Task (NWT). A standalone PBA field test form will consist of only one of these three types of tasks. All students at a specific grade level within a school will take the same task type, however the actual task may vary within a grade. The ELA/literacy PBA standalone field test will include both computer-based and paper-based forms.

Prior to calibration, classical item analysis will be conducted and results evaluated to identify field test items that are potentially flawed or not performing as expected.

The manner by which the standalone PBA field test items will be placed on scale will depend upon whether the results from the field test analyses indicate that ELA/literacy PBA and EOY components can be placed on the same scale.

To support scaling the PBA field test tasks, we propose to use a common item linking design. In this design, some number of PBA tasks from the current year's operational administration will be spiraled among the field test tasks. These tasks will to serve as common item linkage to place the newly developed PBA tasks onto the PARCC operational scale.

Standard item analysis and calibration procedures outlined in section V.C.1. D. will be used to evaluate the performance of the PBA field test items and place them on the PARCC operational scale.

Requirement

V.C.1.F. Retests Scores

Response

Retest Scores

Beginning in Year 2, PARCC will offer retesting opportunities for high school students during regularly scheduled test administration windows for Fall/Winter Block, Spring Traditional, and Spring Block. Summer retesting is not part of the base scope of work. States can work with Pearson separately to add summer retesting for summer 2016 and/or 2017 if needed. Individual PARCC states will decide whether to administer the PBA, or EOY, or both components as the retest opportunity for their students.

Operational sections of both the EOY and PBA components available for retests will include field test or linking items, although they will not be scored. Item parameters obtained during the operational assessment will be used for scoring of the retests. Scoring of retests for students taking both components will be identical to scoring of first-time tests.

However, in cases where students retest in only one of the two components, psychometric and policy issues need to be considered. The driving question is: “Can a score from an earlier administration be combined with a score from another retested component (PBA or EOY) to create an overall combined score?” If so, scoring of the aggregated retest is again straightforward, although the scoring system would need to piece together test components across multiple administrations.

The alternative is to base scoring of the retest only on the component administered during the retest window. Whether scores based only on the PBA or EOY component alone are comparable to scores based on both components collectively depends both on the definition of “comparable” and on the psychometric properties of each component and of the total summative score. The range of possibilities and the psychometric conditions that would produce them are summarized in the following figure.

Possible Conditions for Interpreting PBA or EOY Results as Comparable to Combined PBA and EOY Results

Degree of Comparability	Definition	Psychometric Requirements
Strict comparability	Initial and retest scores have the same expected value and degree of precision for all students. Initial and retest scores are thus truly interchangeable.	▪ Both the PBA and EOY components are unidimensional. ▪ More importantly, the total score composite that aggregates the PBA and EOY components is unidimensional as well. ▪ The two above conditions imply that either the PBA or EOY components provide measurement parallel to that of the composite which combines the two. ▪ Scores based on either the PBA or EOY component are as precise as scores based on both.
Tau-equivalence	Initial and retest scores have the same expected value for all students. However, initial scores, being based on both components, are more precise.	▪ Both the PBA and EOY components are unidimensional. ▪ More importantly, the total score composite that aggregates the PBA and EOY components is unidimensional as well. PBA and EOY items are calibrated with respect to the same ability metric. ▪ Scores based on either the PBA or EOY component alone are sufficiently reliable to report. However, they are less reliable than scores based jointly on both components. ▪ The above conditions imply that either the PBA or EOY components provide measurement that is tau- equivalent to that of the composite which combines the two.
Group mean comparability	Initial and retest scores have the same expected value across all students but not necessarily for each individual student. Total scores for retested students are computed from the component that they were administered as the expected total score given their performance on the observed component. The total score is then estimated as a projection or prediction. Initial and retest scores share a concordance relationship rather than being truly comparable.	▪ While either or both of the PBA and EOY components may be unidimensional, the total score composite is not. PBA and EOY items are likely calibrated on different ability scales. ▪ Expected scores on both the EOY and PBA components, as well as on the total score composite may be unequal for any or all students. ▪ Score reliability may differ between PBA and EOY, and both those reliabilities will certainly differ from that of the total score composite.

Score computation will differ under the comparability conditions outlined above. Under tau-equivalence, the component observed for a given retested student is essentially treated as a representative sample of the total test. Either ability estimates or expected true scores are appropriate scoring methodologies.

Either would be equated (or pre-equated) across forms and across time by verifying that all item parameter estimates are maintained on a common ability metric.

Scoring under the group mean comparability condition is more complicated. Total scores are essentially predicted for each retested student, given the test component that was observed. This prediction is empirical, based on parameters estimated from the entire student population. The total score for a retested student with only an observed EOY component is then estimated as the mean total score across other students at the same level of EOY performance.

Regression methods are typically applied to produce these estimates. Either is population dependent and subject to change as the relationships between the two test components and the composite change. As such, the regression parameters should be re-estimated periodically based on the most current data. .

Requirement

V.C.1.G. Review of Psychometric Analyses

Response Requirements for Section V.C.1.

a) Description of the approach and procedures Offeror will use to complete all the responsibilities/tasks specified in Section V.C.1.

Response

Review of Psychometric Analyses

When all psychometric analyses are completed, ETS will present all results from calibration, scaling and equating of the summative assessments to the content specific Psychometric Analysis Review Committees for approval. The logistical details of these meetings will be the same as the Data Review Committee meetings described in Section V.A.1.J, with the exception of the number of committee members.

There will be one Psychometric Analysis Review Committee for ELA/literacy and another for mathematics. Each Psychometric Analysis Review Committee will consist of four PARCC state representatives and two representatives from the Partnership Manager. Additional attendees will include a psychometrician from Pearson and Measured Progress (the independent auditor), and Pearson program management staff.

ETS psychometric staff will establish the agenda, prepare the necessary materials for the meeting, and record and distribute meeting minutes. Pearson will coordinate the logistics of the meeting and will support the ETS psychometricians in provide all materials for the psychometric review to the Partnership Manager for approval, at least one week prior to the meeting. Pearson program management staff will facilitate necessary aspects of the meetings, logistics, electronic equipment, and other meeting supplies as needed.

V.C.2. Data Forensics for Operational and Retest Assessments

Requirement

Response Requirements for Section V.C.2.

a) Include in base proposal:

i. Description of the approach and procedures Offeror will use to complete all the responsibilities/tasks specified in Section V.C.2.

Deliverables for Section V.C.2.

a) Specifications for Data Forensics at least four (4) weeks before the administration of any assessment.

b) Report summarizing findings of analyses described in the Data Forensics specifications no later than 2 months after the administration of the assessment.

c) Database of proctors/classrooms, schools and districts that were flagged by statistical data forensic methods during the data analyses.

Response

Data Forensics and Validity

Maintaining the validity of test scores is essential in any high-stakes assessment program, and misconduct represents a serious threat to test score validity (e.g., Cizek, 1999; NCME, 2012). When used appropriately, data forensic analyses can serve as an integral component of a wider test security protocol. The results of these data forensic analyses may be instrumental in identifying potential cases of misconduct for further follow-up and investigation.

Using transparent, industry-standard statistical methodologies, Pearson will perform data forensic analyses and look for evidence of anomalous test data, which can sometimes act as an indicator of potential test misconduct. Test misconduct can arise from numerous sources and can take various forms. Some examples of test misconduct include, but are not limited to the following:

- Educators providing students with answers prior to the test administration
- Educators providing students with assistance during the test administration, such as offering hints or answers, or allowing assistive materials to remain accessible
- Educators discouraging historically low-achieving students from taking the test
- Educators changing students' submitted incorrect item responses and replacing them with correct responses
- Test-takers copying item responses from others
- Students sharing items and answers with other test-takers
- Test-takers bringing notes or "cheat sheets" to the test administration

When misconduct occurs it may manifest itself in test data; however, just as the forms of misconduct vary, so do its data manifestations. For this reason, it is necessary to employ several data forensic analysis techniques when investigating for evidence of potential misconduct. No single statistical analysis can adequately identify all possible indicators of potential misconduct.

Pearson proposes a comprehensive data forensic analysis plan. In performing these analyses, Pearson will use established, transparent methods. This proposed forensic analysis plan includes one or more techniques falling under each of the following four categories:

- Longitudinal performance modeling
- Response change analysis
- Response similarity analysis
- Aberrant response pattern detection

Longitudinal Performance Modeling

An observation of excessive student performance gains over time may indicate potential test misconduct (e.g., Jacob & Levitt, 2002; NCME, 2012). A classroom, school, or district (or more generally, unit) that shows evidence of improbably-large performance gains over time may elicit suspicion.

As a strategy to evaluate the reasonableness of observed test performance changes, Pearson proposes using a multinomial regression model to predict test-takers' performance level categories in the current year from their prior test scores.

The outcome variable in the prediction model is the student's observed performance level category at the current time point, and the predictor variable is the student's scale score at the previous time point; therefore, this is characterized as a cumulative logit regression model (Agresti, 1996). From the cumulative logit regression model, we obtain individual probabilities of a given student scoring into each performance level category, conditional on his or her previous test score.

As described in Clark, Skorupski, and Murphy (2013), we treat individual probabilities as examinee-level expected values, then aggregate these and compute the expected count of examinees at performance level j (j = 1, 2, …, J) for unit k by summing independent probabilities for performance level j across all examinees in the unit (i = 1, 2, …, n_k), conditioning on their respective scale scores from the prior time point. Dividing these expected counts by the total number of examinees in the unit yields the expected proportion of examinees at performance level j:

$$E(P_{jk}) = \frac{\sum_{i=1}^{n_k} P_{ij}}{n_k}. \quad (1)$$

The standardized residual for performance level j for unit k is the difference between the observed proportion, P_{jk}, and the expected proportion, $E(P_{jk})$, divided by the standard error:

$$SR_{jk} = \frac{P_{jk} - E(P_{jk})}{SE_{jk}}, \quad (2)$$

where the standard error is equal to the square root of the variance of the mean:

$$SE_{jk} = \sqrt{\frac{E(P_{jk})\left(1 - E(P_{jk})\right)}{n_{jk}}}. \quad (3)$$

Units with large, positive standardized residuals for performance levels corresponding to a level of proficient are indicative of larger-than-predicted proportions of test-takers at these performance levels. An example of the reporting that might occur based on longitudinal performance monitoring is illustrated below for a given class or session. The report includes observed percentages of test-takers at each performance level, predicted percentages at each performance level, and flagging results based upon observed percentages exceeding the agreed-upon threshold value.

Test	Administration	State	District	Site	Class/Session	Students
ELA EOY 10	Spring 2016	XX	XXXXXX	XXXXXX	XXXXXX	XX

Predicted %				Actual %				Flag			
Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	XX.X	*	*	*	*

Response Change Analysis

Tracking changes to submitted responses is a technique that has seen widespread use in large-scale educational testing (e.g., Chute & Niederberger, 2012; New Jersey Department of Education, 2012; Pell, 2012; NCME, 2012). Due to the pervasive nature of paper-and-pencil testing in the history of educational assessment, this technique has often been referred to colloquially as "erasure analysis," but we will use the more general term of response change analysis, which is equally applicable to both paper- and computer-based testing.

For selected-response items such as multiple-choice where there is one key and two or more distractors, there are three possible directions in which response changes can occur: right-to-wrong (RTW), wrong-to-wrong (WTW), and wrong-to-right (WTR). In the context of data forensic analysis, unusually high rates of WTR response changes may elicit suspicion of possible misconduct.

Pearson will record, summarize, and report response changes. Test-takers will be grouped into units, and Pearson will compute the average number of WTR response changes per test form at each unit. An illustrative response change analysis report is shown below. The report is for a particular class and summarizes the means for all item changes, WTR, RTW, and WTW changes. In this case, classrooms found to have a WTR response change average exceeding the agreed-upon threshold would be flagged, with the threshold set at some fixed number of standard deviations above the overall WTR average.

Test	Administration	State	District	Site	Class/Session	Students
ELA EOY 10	Spring 2016	XX	XXXXXX	XXXXXX	XXXXXX	XX

All Items	WTR	RTW	WTW	Z-Score	Z Flag
X.XX	XX.XX	XX.XX	XX.XX	X.XX	*

Response Similarity Analysis

Various forms of misconduct may result in pairs of test-takers having highly similar response patterns. For example, if one test-taker copies answers from another test-taker, this behavior will necessarily inflate the level of similarity in the two test-takers' item response patterns. Incidences where two test-takers select the same response option on a given item is referred to as an identical response. Identical responses are further differentiated into identical correct responses (ICR) when both test-takers choose the correct option and identical incorrect responses (IIR) when both test-takers choose the same distractor.

Numerous response similarity statistics have been developed, including the eight statistics alphabetically labeled *A* through *H* proposed by Angoff (1974), as well as *K* (Holland, 1996), $\overline{K}_1$ and $\overline{K}_2$ (Sotaridona & Meijer, 2002) and S_1 and S_2, (Sotaridona & Meijer, 2003) most notably. All of these techniques are rooted in classical test theory, and with the exception of S_2, all of them draw information from IIR only. In contrast to these techniques, Wollack (1997) developed the omega (ω) statistic, which includes person- and item parameter estimates from the nominal response model. This statistic is computed as

$$\omega = \frac{h_{CS} - E(h_{CS}|\theta_C,\ \mathbf{U}_S, \boldsymbol{\xi})}{\sigma_{h_{CS}-E(h_{CS}|\theta_C, \mathbf{U}_S, \boldsymbol{\xi})}}, \tag{4}$$

where *C* denotes the test-taker in the pair who is being evaluated as the potential copier, *S* denotes the test-taker being evaluated as the potential source, h_{CS} is the observed number of identical item responses between *C* and *S* (including both identical incorrect and identical correct responses), θ_C is the latent achievement trait for *C*, U_S is the observed response vector for *S*, and $\boldsymbol{\xi}$ is the matrix of estimated item parameters (Wollack, 1997).

The omega statistic is computed by subtracting the expected number of identical item responses between test-takers *C* and *S*, which is obtained by conditioning response agreement on *C*'s latent trait, *U*'s observed response vector, and the items' parameters, from the observed number of identical responses. The statistic is standardized by dividing the difference between counts of observed and expected identical responses by their standard deviation. Unlike most of the aforementioned CTT-based response similarity statistics, omega includes information from both identical incorrect as well as identical correct responses, thus increasing the amount of information available from which to draw inferences. By controlling for test-taker latent traits, the potential confounding impact of test-taker achievement is reduced.

In the research literature, omega has been widely studied and generally shown to out-perform other techniques in terms of detection power in head-to-head comparisons (e.g., Wollack, 2006). This finding was confirmed by our own internal research (Clark et. al., 2013). We propose using the omega statistic to evaluate the reasonableness of levels of response pattern similarity among pairs of test-takers.

Based on a review of the research literature and our own findings from an internal investigation, we further propose slight modifications to the flagging procedure:

- First, although this statistic evaluates one examinee as the potential copier and the other as the potential source, we recommend that any such language be removed from reporting, and instead focus flagging on pairs of identified students, with no *C* or *S* label attached to either.
- Second, although this statistic is standardized and is therefore theoretically normally distributed, we propose using an empirically derived flagging procedure based on the observed distribution of values as opposed to applying a theoretically derived flagging criterion, such as 3.00.

An illustrative similarity analysis report for a pair of test-takers is shown below. For each student, the overall raw scores for each test-taker is shown, along with the number of identical correct, incorrect, and total identical responses between them, the number of responses between them that would be expected to be identical, and the resulting Omega statistic. It should be noted that because the number of student pairwise comparisons quickly becomes very large as the cohort over which comparisons are to be made, response similarity analyses may be best conducted only when an irregularity report or a separate forensic analysis suggests some evidence of an anomaly.

Test	Administration	State	District	Site	Class/Session
ELA EOY 10	Spring 2016	XX	XXXXXX	XXXXXX	XXXXXX

Student A		Student B		Identical Responses				
ID	Raw Score	ID	Raw Score	Correct	Incorrect	Total	Expected	Omega
XXXXXXXXX	XX	XXXXXXXXX	XX	XX	XX	XX	X.XX	X.XX

For constructed-response items, Pearson proposes to use our Latent Semantic Analysis (LSA) technology to detect possible plagiarism. Using LSA, we can compare the content of each constructed response against the content of every other constructed response looking for high degrees of similarity. Because LSA provides a semantic representation of language, rather than a syntactic or word-based representation, it allows us to detect potential copying even when students have substituted synonymous words or phrases.

For any cohort of examinees evaluated, Pearson will provide a file highlighting all flagged pairs of examinees, based on either the omega statistic for selected-response items or LSA for constructed-response items. Flagged pairs of show excessive levels of similarity, which may be a possible indicator of non-independent test-taking.

Aberrant Response Pattern Detection

Various forms of misconduct, including but not limited to test-takers referring to “cheat sheets” during the exam, test-takers being exposed to items and answers prior to the test, or educators providing test-takers with assistance during the test, may result in what are referred to as aberrant response patterns. Generally speaking, an aberrant response pattern is simply a set of responses that is statistically unlikely.

As a simple example, a student who achieved a raw score of 5 on a 10-item test by answering the five easiest items correctly has a response pattern that is aligned with expectation. Conversely, had that examinee achieved a score of 5 by answering the five most difficult items correctly, we might conclude that such an outcome is statistically unusual.

These types of patterns are referred to as Guttman patterns and reverse Guttman patterns, respectively (Guttman, 1944). However, most observed response patterns are not perfect Guttman- or reverse Guttman patterns, so person-fit statistics have been developed to assist with measuring the extent to which an observed response pattern aligns with expectation, with expectation being defined in a number of different ways.

A very large number of person-fit statistics have been developed over the past several decades. For example, Karabatsos (2003) compared 36 of these statistics. As is true for response similarity statistics, some person-fit statistics are rooted in CTT, and others are based in IRT. However, much of the research literature has found inconclusive results regarding the use of any one technique or family of techniques. For this reason, Pearson recommends that a collection of person-fit statistics be used.

The modified caution index (*MCI*; Harnisch & Linn, 1981) compares the covariance of a given test-taker's response pattern with two baselines: a Guttman pattern and a reverse Guttman pattern. *MCI* is computed as

$$MCI = \frac{\mathrm{cov}(\mathbf{X}^{*}, \mathbf{p}) - \mathrm{cov}(\mathbf{X}, \mathbf{p})}{\mathrm{cov}(\mathbf{X}^{*}, \mathbf{p}) - \mathrm{cov}(\mathbf{X}', \mathbf{p})}, \qquad (5)$$

where **X** is a given examinee's observed response vector, **X*** is a Guttman response vector, **p** is the vector of observed p-values, and **X'** is a reverse Guttman vector. *MCI* has a lower bound of 0 when **X** is a perfect Guttman pattern and an upper bound of 1 when **X** is a perfect reverse Guttman pattern.

The H^T statistic (Sijtsma, 1986) is computed by

$$H^T = \frac{\sum_{a \neq b} \sigma_{ab}}{\sum_{a \neq b} \sigma_{ab}^{\max}}, \qquad (6)$$

where the observed covariances between all pairs of examinees, indexed *a* and *b*, are summed and divided by the sum of the maximum possible covariances between all examinee pairs. A positive value of H^T is indicative of an observed response vector that is similar to those of other examinees within the data set, and a negative value is indicative of a dissimilar, or aberrant, response pattern.

Finally, the l_z statistic (Drasgow, Levine, & Williams, 1985) is computed as

$$l_z = \frac{l_0 - E(l_0)}{\sqrt{\mathrm{var}(l_0)}}, \qquad (7)$$

where $E(l_0)$ and $\mathrm{var}(l_0)$ are the expectation and variance of l_0, which is the value of the log-likelihood function evaluated at the test-taker's maximum likelihood estimate of θ. Because aberrant response patterns are noted to result in flatter log-likelihood functions, the l_z statistic uses this information to identify test-takers with aberrant response patterns.

Consistent with our recommendations regarding response similarity flagging, we propose that empirically-derived flagging thresholds be used for these person-fit statistics. Individuals flagged as having aberrant response patterns show evidence of unusual test performance, which may or may not be caused by a form of misconduct.

Pearson will provide a file including values for all three person-fit statistics for all test-takers, with flagging outcomes noted. An illustrative report for an individual student is provided below.

Test	Administration	State	District	Site	Class/Session	Student
ELA EOY 10	Spring 2016	XX	XXXXXX	XXXXXX	XXXXXX	XXXXXXXXX

Raw Score	Scale Score	MCI	MCI Flag	HT	HT Flag	LZ	LZ Flag
XX	XXX	X.XX	*	X.XX	*	X.XX	*

As implied by their name, person-fit statistics are predominately focused on identifying data anomalies associated with misconduct at the individual level; however, certain forms of systemic misconduct may result in more widespread incidences of observed aberrant response patterns across a unit.

In addition to the examinee-level results file, we will provide a unit summary file, with units flagged based upon observed rates of response aberrance exceeding an agreed-upon threshold.

Internet and Social Media Monitoring Services

Pearson will collaborate with Caveon Test Security to provide Internet and social media monitoring services proposes. Caveon's team will patrol the Internet, websites, blogs, discussion forums, video archives, social media, document archives, brain dumps, auction sites, media outlets, peer-to-peer servers, etc., for information related to the PARCC Spring, End of Course, small volume block schedule, and summer retake administrations.

The Caveon Web Patrol service addresses risks to test and items posed by illicit discussion, distribution, and sale of test content on the Internet. This service uses a suite of proprietary search methodologies and technology tools, in concert with human expertise. Caveon will generate regular updates that will categorize identified threats by level of actual or potential risk based upon the representations made on the web sites, or actual analysis of the proffered content.

Web sites and Internet extracts are ranked from CLEARED (Lowest risk but are continually monitored for updated content) to SEVERE (Highest risk). Updates contain all needed information, including specific URLs, to quickly evaluate and begin the process of eliminating the threat.

Caveon's Internet and social media monitoring services will include:

- Ongoing intensive Web Patrol monitoring over an eight month period of testing each year, to be determined in conjunction with PARCC.
- Suspected threats will be thoroughly investigated and summary reports will be provided to PARCC through notification emails.

Summary Report

In addition to the aforementioned data files to be provided, Pearson will produce a report documenting the statistical methodologies employed in the data forensic investigation and a summary of relevant findings.

V.C.3. Systems for Data Analysis

Requirement

Response Requirements for Section V.C.3.

a) Description of the approach and procedures Offeror will use to complete all the responsibilities/tasks specified in Section V.C.3.

Deliverables for Section V.C.3.

a) Electronic student data files organized by state, district, school, and grade with the following at a minimum: student raw scores, scale scores, item-level responses (scored and unscored), domain and subscale scores, date administered, form administered, derivative scores (such as growth), and other variables to be named by the Partnership.

b) Agreement that all task and item parameters and other results will be provided in an electronic data file in a format to be designated by PARCC.

c) Flow diagram of how data will travel between system components and be processed and analyzed for psychometric analysis

Response

Systems for Data Analysis

The systems for data analysis related to the psychometric services will be maintained by ETS. Our staff is proficient with a variety of commercially available and open-source psychometric software programs.

For the PARCC Operational Assessments, we plan to continue using the same systems for data analysis that are currently being used for the PARCC field test analyses. We will perform all statistical analyses using publically-available software packages, which will allow for replication by PARCC and independent vendors, as well as ease of transition to future contractors.

Software we believe could be useful to PARCC is described in the figures that follow. We have rank ordered these software packages according to our evaluation of how effective they are likely to be for PARCC work and by their function. On a scale of 1 to 5, a rank of "1" is most likely to support consistent and accurate results— and indicates the software currently in use for the field test.

We will work with PARCC to create an effective solution with any of the software packages listed or other software that is preferred by PARCC and its TAC. We also are willing to work with other commercially available and/or open-source software packages that we have not listed if psychometric properties have been established.

Data Manipulation, Raw Scoring, Classical Item Analysis, and Differential Item Functioning Software			
Proposed Software	**Software Availability**	**Comments**	**Rank**
SAS	Commercially available	SAS is a capable and flexible software package. Because the input is a parameter file, we can publish the program code for IA and DIF in the technical report for easy replication at any time. Because SAS is used in many organizations, it would be relatively easy for PARCC to identify vendors/staff to replicate our work. We propose SAS in performing data manipulation, item scoring, classical item analysis, and DIF.	1
R (base module)	Open source	R (base module) is well-known in research and academic applications, with many software modules making it nearly as flexible as SAS. Because the input is a parameter file, we can publish the program code for item analysis and DIF in the technical report for easy replication at any time. Because R (base module) is used in many organizations, it would be relatively easy to identify staff to replicate our analyses.	2
SPSS	Commercially available	SPSS is similar to SAS in many respects. However, we are aware of no characteristics relevant to this proposal for which it is superior to SAS. It also appears to be used less often in operational settings, which could generate problems as the program is shared from one contractor to another.	3
jMetrik	Open source	A specialized item-analysis package that appears to perform an array of item analysis and DIF procedures well. More analyses are required to investigate suitability for use in the context of high-stakes, large-scale assessments.	4
Iteman	Commercially available	Iteman is a specialized item-analysis package that can handle partial overlap of test forms by creating a criterion score that is based on a subset of items. As	5

Data Manipulation, Raw Scoring, Classical Item Analysis, and Differential Item Functioning Software			
Proposed Software	**Software Availability**	**Comments**	**Rank**
		with jMetrik, more analyses are required to investigate its suitability in the context of high-stakes, large-scale assessments.	

Recommending Software for PARCC. We have rank ordered these software packages according to our evaluation of how effective they are likely to be for PARCC work and by their function, with "1" being our top recommendation and the most likely to support consistent and accurate results.

IRT Scaling Software			
Proposed Software	**Software Availability**	**Comments**	**Rank**
PARSCALE	Commercially available	PARSCALE is likely the most widely used software in operational programs that is capable of calibrating items using the commonly used unidimensional IRT models. Most major testing companies use PARSCALE in at least one program, which speaks to both its quality and flexibility. Also, it will give PARCC the ability to easily transition data from one contractor to the next. PARSCALE has been widely researched, which provides confidence that the results it produces can be relied upon.	1
IRTPRO	Commercially available	IRTPRO is a relatively new software package that can calibrate items using the IRT models supported by PARSCALE, as well as multidimensional models. IRTPRO can also perform multi-group IRT analyses, which could be helpful, were there to be state-to-state variation in item performance due to variations in curricula. However, IRTPRO is relatively new and has had limited use in operational settings. Our view is that IRTPRO is research software that we hope will be ready for operational use soon. However, at this time, more analyses are required to investigate its suitability in the context of high-stakes, large-scale assessments.	2
WINSTEPS	Commercially available	WINSTEPS is a well-known and widely-used software package for implementing Rasch models. Other models outside the Rasch family cannot be evaluated using it.	3
ICL	Open source	ICL can calibrate items using most of the IRT models supported by PARSCALE, including 1-, 2- and 3-parameter logistic models, partial credit, and generalized partial credit. ICL can also perform multigroup IRT analyses, which could be helpful were there to be state-to-state variation in item performance due to variations in curricula. Our view is that ICL is research software, and more analyses are required to investigate it suitability in the context of high-stakes, large-scale assessments.	4
R (ltm)	Open source	R (ltm) can calibrate items using most of the IRT models supported by PARSCALE, including 1-, 2- and 3-parameter logistic models, as well as the generalized partial-credit model. Our view is that R (module ltm) is research software that lacks sufficient evidence to support	5

IRT Scaling Software			
Proposed Software	**Software Availability**	**Comments**	**Rank**
		its operational use. At this time, more analyses are required to investigate it suitability in the context of high-stakes, large-scale assessments.	

Recommending Software for PARCC. We have rank ordered these software packages according to our evaluation of how effective they are likely to be for PARCC work and by their function, with a "1" representing the software most likely to support consistent and accurate results.

Additional Psychometric Software				
Function	**Proposed Software**	**Software Availability**	**Comments**	**Rank**
IRT Equating	STUIRT	Open source	STUIRT conducts IRT scale transformations for mixed dichotomously and polytomously scored tests. It supports the unidimensional IRT models supported by PARSCALE. This software was developed at the University of Iowa, and it has been used in statewide testing programs, including Maine, Maryland, Michigan, and Oklahoma.	1
	IRTEQ	Open source	IRTEQ conducts IRT scale transformations for mixed dichotomously and polytomously scored tests. This software appears to be as capable as STUIRT, but it is not preferred due to STUIRT's record of success.	2
	Plink	Open source	Plink conducts IRT scale transformations for mixed dichotomously and polytomously scored items. Its properties are not well-known.	3

Recommending Software for PARCC

We have rank ordered these software packages according to our evaluation of how effective they are likely to be for PARCC work and by their function, with a rank of "1" as the most likely to support consistent and accurate results. ETS will make available any software code and/or control files, as applicable, associated with any of the PARCC-approved analysis software programs as part of the technical documentation. We will keep all the relevant documents and electronic files (including software code) in a format approved by PARCC.

V.C.4 Technology and Data Requirements

Requirement

V.C.4.A. PARCC Data Management and Reporting Platforms

Response

PARCC requires a vendor familiar with the complexity of the many systems managed for such a large-scale assessment. As we describe in section V.B.1, the technology philosophy and systems architecture in our solution have been designed to support states in administering next-generation assessments in a way that empowers PARCC's long-term sustainability.

Pearson will use the data management, reporting, and analytics system being created by the PARCC Technology Bundle vendor to fulfill the activities and deliverables required under this RFP. We will coordinate with this vendor starting early in the contract to coordinate planning regarding data handoffs.

Under this contract, Pearson will use the PARCC Data Management and Reporting system to load the results of statistical analysis data with the student data records to be integrated together.

Pearson works with providers to create secure hosting environments that protect network assets and information from unauthorized access or operations disruptions. We maintain a Security Policy and Requirements document outlining strict procedures for the physical security of hosting facilities.

Protection from network-based threats is as important as physical security. Pearson engineers, in conjunction with data center staff, deploy, manage, and monitor the security of Pearson systems. We employ a variety of security technologies and tools in the computing and network environments. External penetration testing and security scanning are routinely performed to verify that our systems are adequately hardened and protected from security threats.

Requirement

V.C.4.B. Interoperability Requirements

Response

As stated in section V.B.1.E, Pearson is committed to industry interoperability standards and works closely with standards organizations to develop and enhance data and content standards. We will draw upon extensive in-house experience to align to industry best practices such as AIF, CEDS, and SIF. For example, Wayne Ostler, Vice President of Digital Content and Measurement Systems; Jason Craft, Principal Software Engineer; and Michelle Richard, Manager Content Encoding and Transformation services; all actively participate on the APIP Working Group (known as the APMG) and are part of the management team that will provide executive leadership and management to the project.

Pearson will use the PARCC schemas and coding guidelines described in the PARCC Item Development Technical Guide (Attachment L) and will work with Partnership representatives to update these schemas as necessary each year based on the Partnership's needs, including allowance for item developer innovation and the results of Research Studies to be conducted by Pearson under this contract. Pearson will update our practices accordingly, based on PARCC approval.

Requirement

V.C.4.C. Data Privacy and Security

Response

Data privacy and security play critical roles in the assessment process. Whether we refer to student demographic information or assessment content, Pearson acknowledges the need for privacy and security and will employ security protocols and design features that meet or exceed PARCC security needs for data privacy and security. This includes the appropriate use of encryption, identity management, controlled data access, and so on. Pearson will comply with federal laws and Partnership policies for data privacy and security.

Section V.B.2.B of our response includes detailed plans of Pearson's standard security and technology policy. The section includes details on the following:

- Overview of Pearson's commitment to security and our security awareness program details.
- Overview of Logical Security controls show we have **redundant** ways to protect PARCC assessments and data.
- The Servicer and System scalability explains our **auditable security**, which includes logging, monitoring, and auditing of PARCC assessments, data, and data access.
- Data Security/Encryption, Secure Access Controls, and Overview of Physical Security Controls explain how we employ identity management and data access are employed from a **security protocols** perspective.

Requirement

V.C.4.D. Technical Integration Requirements

Response

An electronic student data file will be uploaded into PARCC's Data Management and Reporting System. As required by the RFP, the file will be organized by state, district, school, and grade and will at a minimum include student raw scores, scale scores, item-level responses (scored and unscored), domain, and subscale scores.

Requirement

Response Requirements for V.C.4.

a) Description of the approach and procedures Offeror will use to complete all the responsibilities/tasks specified in Section V.C.4.

Deliverables for Section V.C.4.

a) Privacy and security plans

b) Electronic student data files uploaded into PARCC's Data Management and Reporting System

Response

Pearson will work with PARCC and Partnership vendors to perform all responsibilities and tasks specified in Section V.C.4 by following required policies and procedures.

V.C.5. Research Studies

Requirement

V.C.5.A. PARCC research conducted under PARCC Field Test Contracts

1. Quality of Items/Stimuli Study
2. Accommodations and Accessibility Study 1: Analyses of Field Test Observations and Psychometric Data
3. Accommodations and Accessibility Study 2: Validity and Accuracy of Scoring for PCR Items
4. Accommodations and Accessibility Study 3: Read Aloud Accommodation Study
5. Mode and Device Comparability Study

V.C.5.B. Additional Psychometric Research to be conducted with Operational Assessment Data

1. Study 1: Comparability of Assessment Results
2. Study 2: Test Administration Mode and Device
3. Study 3: External Validity of Read-Aloud Text-to-Speech Accommodation
4. Study 4: Accessibility of New Items/Functionalities and Use of New Devices
5. Study 5: International Benchmarking Study

Response Requirements for V.C.5.

a) Description of the approach and procedures proposed to complete all the responsibilities/tasks specified in Section V.C.5
b) Description of the nature of coordination required with other PARCC Contractors and with the third party organizations in conducting each study.
c) Description of general approach and the rationale for the:
    i. study design and data analyses including (but not limited to) approach and rationale for selection of items and tasks to be included in the studies, and data collection tools and methods if and where applicable
    ii. sampling frameworks (including sample size) and procedures for recruiting and securing required sample sizes if sampling is required
d) Plans for mitigating the negative impact of missing data, where appropriate.
e) Description of quality control procedures to ensure accuracy in data processing

Deliverables for Section V.C.5.

a) Draft for each study plan by May 2, 2014. The plan shall include timelines, study design, sampling specifications, and data analysis methods
b) The Contractor shall present the study plans to the PARCC Technical Advisory Committee (TAC), one or more PARCC Operational Working Groups (OWG), and/or other expert reviewers for feedback. The Contractor shall incorporate such feedback.
c) Final Study Plan for each study in 4 weeks following PARCC feedback on draft plan.
d) Draft data collection instruments, survey and/or data coding schemas 8 weeks in advance of the start of data collection
    i. Contractor shall provide all data collection instruments (e.g., interview protocols, observation protocols, surveys) and coding schemas to the Partnership Manager in draft form. Each data collection instrument and coding schema will be reviewed by the TAG, one or more PARCC OWGs, and/or other expert reviewers. The Contractor shall revise the data collection instruments and coding schema accordingly, prior to use in the study.
e) Final data collection instruments, survey and/or data coding schemas in 4 weeks following PARCC feedback on drafts

f) Draft study reports within six weeks of the completion of data collection or a date mutually-agreed upon by the Partnership and the Contractor.
    i. Draft reports for each study will be reviewed by the TAG, one or more PARCC OWGs, and/or other expert reviewers for feedback.
    ii. Draft reports shall indicate the Principal Investigator, data analysts and Contractor staff who reviewed and approved the submission of the draft report
g) Final study report within two weeks after the feedback for the draft report is provided or a date mutually-agreed upon by the Partnership and the Contractor
    i. The final report must include the theoretical framework and design rationale that cites relevant, peer-reviewed published work and, when appropriate, unpublished technical reports; how the study addressed relevant standards in the Standards for Educational and Psychological Testing (1999); and how the results contribute to the body of evidence to support the valid interpretation of scores.
    ii. The final report shall include an executive summary of results, and specific recommendations of action
    iii. Final reports shall indicate the Principal Investigator, data analysts and Contractor staff who reviewed and approved the submission of the final report
h) An action plan based on recommendations in the final study report
    i. The action plan shall indicate responsibilities of each party involved along with a timeline for each action
    ii. The Contractor shall manage the action plan and provide update reports for actions that require Contractor follow up or involvement based on a schedule proposed by the Contractor and mutually-agreed upon by the Partnership and the Contractor
i) All data (raw, or scored, or coded or processed) collected and processed for each study in 2 weeks following the delivery of the final study report in a digital format proposed by the Contractor and approved by PARCC.

Response

PARCC has identified five research studies to be conducted with Operational Assessment Data. In many cases, the studies continue investigations that will be carried out as part of the field test administration, and the results of these studies will inform various decisions related to the scoring and scaling of PARCC assessments.

Pearson will work with our delivery collaborator, ETS, to carry out these five research studies. ETS researchers will be involved in all five of the studies and will lead three of them. Researchers from Pearson will play prominent roles in two of the studies, capitalizing on their expertise and experience in particular topics. The specific roles of ETS and Pearson researchers in the five studies are as follows:

- Study 1: Comparability of Assessment Results—ETS to lead
- Study 2: Test Administration Mode and Device—ETS and Pearson to lead jointly
- Study 3: External Validity of Read-Aloud Text-to-Speech Accommodation—ETS to lead
- Study 4: Accessibility of New Items/Functionalities and Use of New Devices—Pearson to lead, ETS to contribute
- Study 5: International Benchmarking Study—ETS to lead

Detailed descriptions of the approach and procedures proposed to complete all the responsibilities/tasks for these five studies are provided in the sections that follow.

Study 1: Comparability of Assessment Results

Study Overview

PARCC Summative Assessments are comprised of PBA and EOY components and will be administered in multiple forms, on many occasions, and across multiple years. To properly compare performance across students, schools and states, scores obtained from different test forms need to be directly comparable. For performance trends to be accurately determined and interpreted over time, score scales must be properly maintained so that scores are comparable across years of test administrations.

The challenges of maintaining score comparability across forms and time are significant in the fluid environment that will likely characterize operational testing. Test administration policies and procedures may differ across states and change over time. Retesting and accommodation policies and administration schedules are just a few examples of these differences. The tested population may also change over time, both as a consequence of PARCC's theory of action and due to various states joining or exiting the consortium. Item development practice and test scoring methodology are also likely to improve over time, benefitting the testing program in general but challenging score comparability nonetheless.

The specific research questions posed by the above challenges include:

1. Are subject-level scaled scores comparable to the point of interchangeability across different forms and across years?
2. What level of comparability exists for domain scores reported at the claim and sub-claim level across years and across forms?
3. Would changes in state participation levels from one year to the next impact item calibration and equating?
4. Can summative assessment results for states administering retests be properly compared to summative assessment results for states not administering retests?
5. Can results obtained from retests based on PBA-only, EOY-only, and both PBA and EOY administrations be reported on the same metric?
6. Can results obtained from the high school block-schedule and end-of-year administrations be reported on the same metric?

Questions 4 and 5 on this list are addressed in Section V.C.1.F of this proposal. Questions 1-3 and Question 6 are discussed below.

Q1: Comparability of subject-level scaled scores between forms and across years

For test scores to be comparable across different forms and time, at least two conditions must be true. The first is that the different (or alternate) test forms are substantively equivalent, meaning that each measures the same construct(s) in the same way. This is done both by defining explicit and detailed test specifications and by verifying that each newly assembled form sufficiently meets those specifications. The second condition is that the equating methodology operationally employed to adjust away subtle and inevitable differences in form difficulty is properly designed and correctly applied. Both of these conditions are required, with neither alone sufficient to provide score comparability.

The most direct evaluation of score comparability is to administer each of several forms to randomly equivalent student samples. Differences in scaled score distributions can then be directly attributed to instability either in form specifications and assembly or in equating methodology or practice. Administering forms developed, assembled, and equated at different points in time to equivalent groups similarly directly evaluates the stability and comparability of scores across administration cycles. We therefore propose that the year 4 administration will spiral in test forms from year 1 or from a series of earlier years. For added strength, the design should include several forms from each point in the program's history. Administering each form in the design to randomly equivalent groups of 3,000-5,000 students will provide a powerful test of score comparability and stability.

Statistical evaluation of results will begin with comparisons of scale score means, variances and distributions. Because each test form in the study had been independently equated to the base score scale, score distributions across forms should differ only by chance, or to the extent that the randomly equivalent groups would be expected to differ. Differences beyond the level of chance would be attributed to score instability. Statistical analyses of score distribution differences will proceed from several directions:

- Comparison of mean differences
- Tests of the hypothesis that all observed distributions are samples from the same population
- Re-equating the year 1 form based on the year 4 sample and comparing that to the initial year 1 equating. Differences would be judged relative to the criterion termed by Dorans and Feigenbaum (1994) as the equating "difference that matters."

IRT methodology will allow the focus to narrow from test scores to individual items by determining the extent to which item parameters estimated and linked to the base ability scale in year 1 remained predictive of student responses in year 4. A number of goodness-of-fit indices are appropriate for this comparison, with each comparing observed response from year 4 students to expectations based on item parameters estimated and scaled from earlier student samples (Donoghue & Isham, 1998; Glas, 2000; Smith, Wang, Wingersky & Zhao, 2002). Careful examination of these results may reveal that certain item types or content areas have remained more stable than have others.

In addition, PARCC may want to conduct elements of the study outlined above prior to year 4. It may in fact be prudent to directly monitor score stability on an ongoing basis by routinely including in each administration year items and test forms calibrated and equated in past years. Early identification of sources of score instability may inform changes in methodology or practice that are able improve score comparability going forward. .

Q2: Comparability of domain scores

Domain scores determined at the claim and sub-claim level will be based on the same IRT parameter estimates as are subject-level scores, allowing them to be equated across forms and administration years. As such, the same analyses outlined above to evaluate the comparability of subject-level scores can be applied to domain scores as well. However, the analyses will be weakened by the fact of domain scores being based on fewer items than subject-level scores. For the same reason, comparability of domain scores is likely to be more difficult to achieve. Scores are always susceptible to idiosyncratic differences in the sampling of items that comprise a form. These differences are more inclined to cancel one another and "wash out" for scores based on more items.

Q3: Impact of state participation rates on item calibration and score equating

Current plans call for the PARCC assessments to be administered across 13 states and the District of Columbia. However, this number may change as the system matures and states choose to enter or exit the consortium. The concern is that such changes to the tested population may impact score comparability across time by affecting the item parameter estimates on which scores are based.

Item parameter estimates are certainly population dependent to some extent. Differences in population composition in terms of student demographics or in ways the curriculum is implemented in can subtly change the observed dimensionality of the test. This in turn can change item parameter estimates and so impact scale linking and score comparability. This study is intended to evaluate the extent of this impact. The simplest and most direct approach is as follows:

1. Using year 1 operational data, calibrate the item parameters with data from all states.
2. Remove the data of one state or a cluster of states that have similar characteristics, and repeat the calibration.
3. Calibrate all items from the reduced data set and link these calibrations to those estimated from all data. This linking emulates what would happen should the calibration population change across administration years.
4. Compute scaled scores for all students based on both the full- and reduced-data calibrations to examine impact on score comparability. Many of the comparative analyses described for use with the two score comparability studies above remain relevant here as well. .
5. Steps 2) through 4) can be repeated, removing other states or clusters of states.

If important differences are found, it is recommended that operational practice include definition of a “reference population” described in terms of its demographic characteristics. Calibration samples would then be routinely post-stratified to agree with the reference population, therefore mitigating the impact of any changes in the composition of the student population.

Q6: Comparability between Block-schedule and Traditional Year

Although the tests administered to block- and traditionally-scheduled schools may be drawn from the same bank of items, there is no assurance that all items will perform identically in the two groups. However, determining whether this is the case is made difficult by the fact that the block- and traditionally-scheduled student groups cannot be assumed equivalent. Comparability of item performance and corresponding test scores will therefore be investigated through two distinct approaches, each making different assumptions and evaluating different aspects of comparability. These analyses directly parallel those proposed for evaluating comparability across math sequences in section V.C.1.D.2.

The primary driver of score equivalence, and so the primary question investigated here, is whether item parameters estimated from traditionally-scheduled students can be properly used to score the tests administered to block-scheduled students. If that is found to be the case, a single set of operational processes could support both traditional- and block-scheduled tests. If it is not true, alternatives for operationally supporting the block-scheduled tests would need to be determined and implemented. .

Approach 1: Item fit analyses. The full student sample will presumably be dominated by traditionally-scheduled schools. Item parameters estimated from this sample are appropriate for scoring block-scheduled students to the extent that they “fit” or are predictive of responses in this subgroup. However, to test the fit of any given item it is necessary to assume that most other items fit acceptably. Most item fit indices work by comparing the response observed for a student with that expected under the presumed IRT model. That expectation is a function both of the parameters for the item in question and the ability estimate of the student. The latter is a function of the IRT parameters for all items in the student’s test. Consider the case where all items misfit in the block-scheduled subgroup by being estimated as too difficult. This shift in item difficulty would be absorbed into the corresponding ability estimates, making each student appear to be too able. The combination of misfit item parameters and biased ability estimates would result in predictions that closely matched the observed responses. Item fit analyses are therefore seen as capable of identifying non-pervasive or idiosyncratic patterns of misfit. The same item fit measures listed above in Study 1 are appropriate for use here as well. Checking the fit of item parameters estimated in year 1 students for predicting the responses of year 4 students is largely the same problem as checking fit of item parameters estimated from traditionally-scheduled schools for predicting the response of block-scheduled students. .

Approach 2: Separate calibration. This analysis will conduct separate calibrations of the items comprising the block-scheduled test forms. These calibrations would then need to be linked to the traditionally-scheduled school ability metric that underlies test scoring and equating. Linking would again need to assume that at least some items perform equivalently in the two student groups and so can serve as anchors. Each block-scheduled student would then be scored from two distinct sets of item parameter estimates. The first is that estimated from the total sample, which again is dominated by traditional schools. The second is that estimated from the block-scheduled students alone and then linked to the total sample estimates. The comparability of the two sets of test scores would then be assessed, again using most of the same analyses proposed for Study 1, above.

Coordination with Other Partners

From planning to final results, we will work with PARCC throughout to implement the most feasible and informative study. The Comparability of Assessment Results study will not require extensive coordination with other PARCC Contractors or third party organizations, as ETS and Pearson will have all operational data necessary to conduct the study.

Sampling and Missing Data Considerations

In the studies described above, data from the administration of operational test forms will be used and spiraled across all examinees. As such, there is no need for separate sampling of either items or tasks, nor of students. In addition, the large samples and the planned analyses do not require special handling of missing data.

Quality Control Procedures

The processes and procedures will be reviewed both internally by other staff members from the Pearson and ETS teams and externally by the PARCC TAC. Quantitative data and analysis results (e.g., item-level statistics, reliability and validity estimates, etc.) will be verified by multiple researchers. Reports will undergo a multi-stage iterative review process including internal review among team members, external review, PARCC review, and final submission.

References

Dorans, N.J. & Feigenbaum, M.D. (1994). Equating issues engendered by changes to the new SAT and PSAT/NMSQT. In I.M. Lawrence, N.J. Dorans, M.D. Feigenbaum, N.J. Feryok, A.P. Schmitt, & N.K. Wright (Eds.), *Technical issues related to the introduction of the new SAT and PSAT/NMSQT* (ETS Research Memorandum No. RM-94-10). Princeton, NJ: Educational Testing Service.

Study 2: Test Administration Mode and Device

Experience and Capabilities

Pearson and ETS have successfully worked together with PARCC to plan for and conduct mode and device comparability studies as part of the field test administration so have a unique perspective on the factors that might influence comparability for PARCC assessments. In addition, both Pearson and ETS have conducted mode comparability research with large-scale assessment clients over the past 10 years as states have explored and implemented online testing. Specifically, both companies have extensive experience conducting mode comparability studies for high-stakes, large-scale K–12 assessment programs. ETS has also conducted numerous studies investigating various aspects of digitally delivered assessments, in both K–12 and non-K–12 testing environments.

More recently, both organizations have conducted research examining device comparability. Pearson, specifically, has an on-going program of research around device comparability and has developed solutions and policy guidance that will allow states to offer CBTs to the widest possible set of devices without sacrificing security, fairness, or validity. An early phase of research in this area looked at comparability of scores across test-takers using netbooks with screen sizes of either 10.1 or 11.6 inches and students using the 14- to 21-inch screens common on desktop and laptop computers (Keng, Kong, & Bleil, 2011). More recent research has focused on the use of tablets for CBT. An initial qualitative study was conducted within a single state using a tablet with and without an external keyboard, a range of item/task types, and three grade levels (Strain-Seymour, Craft, Davis, & Elbom, 2013). A second qualitative study evaluated a wider range of devices and item types and involved four states: Florida, Texas, Maryland, and Virginia (Davis, Strain-Seymour, & Gay, 2013). The next phase of research (with results expected in early spring 2014) extends the finding of our previous research to quantitatively compare student writing when essays are generated via a laptop, a tablet, and a tablet with an external keyboard.

Study Overview

The need to evaluate the comparability of test scores when an assessment is delivered via both paper and computer is addressed in the professional testing standards (APA, 1986; AERA, APA, NCME, 1999, Standard 4.10). Comparability research over the last quarter century has largely focused on differences between paper-and-pencil testing (PPT) and computer-based testing (CBT); however, with an ever expanding range of digital devices appearing in the classroom (e.g. tablets, netbooks, hybrid convertibles, etc.), the definition of CBT has become much broader. As such, comparability across device type within the CBT mode has also become an important consideration.

PARCC's ultimate goal is digital delivery of the ELA/literacy and mathematics assessments using the widest variety of devices that will support interchangeable scores. As stated in the RFP, these include desktop computers, laptops, and tablets.

The proposed comparability study will compare performance at the item and test level for students testing on computers and those testing on touch-screen tablets. In addition, because PPT will also be provided as an option for schools where technology infrastructure is not ready for digital delivery, the study will also address the comparability between results based on CBT and PPT at both the item- (where applicable) and test-level. The study will evaluate the degree to which scores can be reported on a single scale versus the degree to which separate scales must be built for different modes and/or devices. It should be noted that the degree of comparability across CBT and PPT may be limited by the inclusion of technology enhanced items (TEIs) in the CBT blueprint (Luecht & Camera, 2011).

The degree of comparability across modes and devices will be an evolving issue throughout the life of the PARCC program. Inevitably the relationship between the two modes will change over the years as the proportion of TEIs on the PARCC CBT blueprints increases and the types of TEIs move from more basic interactions to more complex simulations and gaming-like experiences. In addition, student familiarity with and use of various devices in an academic context will also evolve as the digital classroom becomes a reality for more and more students. In essence, the degree of mode and device comparability is a target that is likely to continuously move as different cohorts of students are studied and salient factors within the PARCC blueprints evolve.

This study will be jointly led by Pearson and ETS researchers with expertise and experience in conducting research in mode and device comparability. Pearson's and ETS's experiences with large-scale dual-mode assessments and on-going research in device comparability position us to meet PARCC's needs in the following areas:

- Evaluating to what degree item-level and test-level scores obtained from CBT and PPT administrations are comparable;
- Evaluating to what degree item-level and test-level scores obtained from various devices are comparable;
- Addressing to what degree comparability can be obtained through scaling items onto a single metric, linking or concordance.

Study Methods

Pearson and ETS propose to conduct both mode (PPT vs. CBT) and device (computer vs. tablet) comparability research in years 1 and 3 of the program for all grades/courses within the "Traditional Year" test administration window. This includes the full summative test (PBA and EOY) for the following assessments:

- Grades 3-11 ELA/literacy
- Grades 3-8 Mathematics
- Algebra I, Geometry, and Algebra II
- Integrated Mathematics I, II, and III

We propose a quasi-experimental approach such as the Matched Samples Comparability Analysis (MSCA) (Way, Um, Lin, & McClarty, 2007; Way, Davis, & Fitzpatrick, 2006) to create randomly equivalent matched groups based on natural participation patterns within the operational test. This approach avoids the need for special sampling or assignment of students to condition. It does assume, however, that a sufficient volume of students will test per form within each mode or device condition to support meaningful statistical interpretation of results at both the item and test level. It further assumes that students will participate in the PARCC test administrations using the mode or device that matches what they use in the course of instruction within their classrooms.

During testing, the software will capture and pass back information about the device a student used during testing. This will form the basis for the initial grouping of students into device conditions. The MSCA approach will then apply bootstrap resampling techniques to match students across PPT, computer, and tablet conditions on a composite variable that reflects a set variables related to students' academic proficiency and demographics. Selecting the matching variables to use in the composite will require different approaches in the first year than in later years of the studies. In later years, students' academic proficiency can be represented via their prior year scores on the PARCC test administrations. However, in the first year, prior year scores will reflect a variety of different state assessment scores as well as student performance on the PARCC field test. This will mean that students will have different measures of academic proficiency which cannot be simply aggregated together. While this can be addressed using a stratified MSCA approach where students are matched only to other students within their same state, the relationship between PARCC scores and state assessment scores will vary across states. Pearson and ETS will work with PARCC to identify the best approach to mitigate these issues within the first year study. Separate but similar considerations will include discussions of how to identify appropriate matching variables for students in grade 3 (where there are no previous test scores) as well as how to incorporate new states which might join the Partnership into the research design.

Both test level and item level analyses will be conducted using data from the matched samples. At the test level, comparisons will be made of the factor structure, reliability, difficulty, and score distribution of test scores (overall and by demographic and state subgroups) resulting from different modes and devices. At the item level, comparisons will be made (where applicable) of item p-values and means and item response theory (IRT) difficulties across different modes and devices. Additionally, Mantel-Haenszel (MH) differential item functioning (DIF) analysis will be conducted using mode or device as the classification variable to further evaluate differential item performance.

As PARCC considers the allowance of additional devices for testing it will be important to evaluate the results of qualitative research conducted under *Study 4: Accessibility of New Items/Functionalities and Use of New Devices* to determine whether new devices can be expected to function similarly to existing devices or whether a separate device grouping would be necessary. Pearson and ETS researchers conducting the quantitative studies will work closely with those conducting the qualitative studies to inform these decisions.

Following the creation of randomly equivalent groups through MSCA, we propose the use of a comprehensive framework for evaluating the comparability of assessment results across the different mode and device conditions. This framework not only evaluates test score comparability, but also examines comparability at the item/task level and takes into account important psychometric properties, such as reliability test structure. More specifically, the proposed framework includes the following components:

Item/Task-Level Comparability

- Do the individual items/tasks perform similarly and rank order similarly across different devices?
- For items which appear in both CBT and PPT modes, do the individual items/tasks perform similarly and rank order similarly across different modes?

Test-Level Comparability

- Would students receive similar scale scores and be consistently classified into performance levels across different modes and devices?
- Are the psychometric properties of the test scores (e.g. factor structure, reliability, difficulty, and score distribution) similar across different modes and devices?

By answering these questions, this framework can evaluate the degree to which item/task-level statistics and test-level scores obtained from the different modes and devices are comparable. It can also help identify features of the devices that interact with item or task properties and student characteristics. Specifically, these study results could:

- Inform changes to future item development, content formatting, or user interface functioning
- Inform changes to future policies around device or peripheral inclusion or requirements
- Confirm use of the same scale across modes and devices or the development of separate scales (if needed) to help facilitate score interchangeability across modes or devices

Different pieces of the comparability framework might inform different actions. For example, item/task-level analyses might best inform future actions around item development, formatting, or device peripheral policies. Similarly, a decision to build separate scales for different modes or devices might best be informed by test-level information. Recommendations about using a single scale or building separate scales for different modes or devices should consider the following three criteria:

1. Statistical significance of differences between test scores across modes or devices
2. Practical significance of differences between test scores across modes or devices
3. Consistency of classification into performance level based on test scores across modes or devices

The mode and device comparability studies will be conducted during years 1 and 3 of the program . This allows for the degree of comparability across modes and devices to be re-evaluated throughout the life of the PARCC program as students gain experience in the CCSS (opportunity to learn), the proportion of TEIs on the PARCC CBT blueprints increases, and the types of TEIs move from more basic interactions to more complex simulations and gaming-like experiences.

Quality Control Procedures

Processes and procedures will be reviewed both internally by other staff members from the Pearson and ETS teams and by the PARCC research staff and their TAC. Quantitative data and analysis results (e.g., item-level statistics, reliability and validity estimates, etc.) will be verified by multiple researchers. Reports will undergo a multi-stage iterative review process including internal review among team members, external review, PARCC review, and final submission.

Deliverables

Pearson and ETS will provide a draft study plan including timelines, study design, and data analysis methods (no sampling will be needed) and will provide a final study plan within 4 weeks following feedback from PARCC on the draft plan. The date for the draft study plan will be mutually agreed upon by Pearson and PARCC during final contract negotiations once contract start date has been determined. In addition, Pearson and ETS will present this plan to the PARCC Technical Advisory Committee (TAC), one or more PARCC Operational Working Groups (OWG), and/or other expert reviewers as requested by PARCC.

If separate scales for different modes or devices are recommended, Pearson and ETS will prioritize building these scales for use in operational scoring and reporting. Pearson and ETS will work closely with PARCC designated representatives to recommend actions and obtain approval based on study results. As these approvals will likely be needed in a very tight timeframe, Pearson and ETS will work with PARCC to determine in advance exactly what materials, analysis results, and other evidence would be needed to support the approval process.

Following the completion of data collection and study analyses, Pearson and ETS will draft a study report summarizing results from all grades/courses with the mode and device comparability study. This will occur within six weeks of the completion of data collection or at a date mutually agreed upon by PARCC and Pearson/ETS. This report will include both item-level (where applicable) and test-level analyses and will be made available for review by the TAC, one or more PARCC OWGs, and or other expert reviewers. Additionally the report will indicate the Principal investigator, data analysts, and contractor staff who reviewed and approved the submission of the draft report. A final report will be provided within two weeks after feedback from PARCC (or at a date mutually agreed upon by PARCC and Pearson/ETS).

The final report will include the theoretical framework and design rationale and cite relevant peer-reviewed published work and, when appropriate, unpublished technical reports. Additionally, the final report will include an executive summary as well as specific recommendations of action along with a timeframe for each action. Following acceptance of the recommendations by PARCC, Pearson and ETS will draft and manage an action plan for implementing the recommendations. Within two weeks of the delivery of the final study report, Pearson and ETS will provide all study data (raw, scored, coded, or processed) in a mutually agreed upon digital format to PARCC.

Study 3: External Validity of Read-Aloud/Text-to-Speech Accommodation

Background and Purpose

Predictive validity (also referred to as criterion-related validity, criterion-related evidence, predictive evidence, or evidence based on relationships to other variables) is a key part of the validity argument that supports the claims made about test score uses and interpretations (see *Standards for Educational and Psychological Testing,* AERA, APA, NCME (1999) pp. 14–15). Because PARCC scores will be used to measure college and career readiness, it is essential that the predictive value of PARCC scores apply to both students taking accommodated test forms and students taking non-accommodated test forms.

Study 3 will examine the external validity of the read-aloud/text-to-speech accommodation offered by PARCC. This study, which will be led by ETS, will compare predictive validity for:

1. general population students without an accommodation;
2. students who need the text-to-speech (TTS) accommodation and are given this accommodation; and
3. students who need the TTS accommodation but are not given this accommodation.

The study is designed to provide evidence to determine if the TTS accommodation removes construct-irrelevant barriers that prevent students from demonstrating their college readiness as measured by the PARCC assessment.

ETS researchers have significant experience in conducting predictive validity studies for admissions tests (see, e.g., Braun, Ragosta, & Kaplan, 1986; Bridgeman & Lakin, 2012; Cahalan-Laitusis, Mandinach, & Camara, 2002, Jones & Ragosta, 1982), experimentally designed research studies investigating testing accommodations for both state assessments and admissions tests (see, e.g., Bridgeman, Laitusis, & Cline, 2007; Laitusis, 2010; Mandinach, Bridgeman, Cahalan (Laitusis), & Trapani, 2005), and studies of the impact of testing accommodations using operational test data (see, e.g., Buzick & Stone, 2011; Laitusis, 2010; and Stone, Cook, Laitusis, & Cline, 2010). ETS conducted the largest repeated-measures study ever conducted on the read-aloud accommodation in 2005 with more than 900 students with learning disabilities and 1,100 students without disabilities. In addition, ETS researchers recently completed a detailed literature review of the read-aloud accommodation for the Smarter Balanced Assessment Consortium (Laitusis, Buzick, Stone, Hansen, & Hakkinen, 2012) and an ETS-funded meta-analysis of the read-aloud accommodation (Buzick & Stone, 2013).

In summary, the ETS team is very well-positioned to execute the proposed study based on its experience conducting research on testing accommodations using a wide variety of methodologies, including large experimentally designed studies (between- and within-subjects designs), qualitative studies using cognitive labs, and studies using operational test data (e.g., differential item functioning, predictive validity studies). We will work with PARCC to refine the proposed design as needed.

PARCC requests evidence to answer two research questions:

1. Does performance on PARCC assessments predict freshman college students' course performance for
   - Group A: general population students
   - Group B: students who need the text-to-speech accommodation and are given this accommodation
   - Group C: students who need the text-to-speech accommodation but are not given this accommodation?
2. How does the strength of the relations between performance on PARCC assessments and freshman course performance compare for the three groups mentioned above?

Possible Methods and Requirements

These research questions can be evaluated under several data collection and analysis designs. Due to the need to test Group C (students requiring the accommodation who will not receive it) for the purposes of this study, an experimental design is required. We recommend an independent-groups (alternatively, between-subjects) design, which is common for predictive validity studies. This design will require random assignment of students requiring the accommodation to either Group B or Group C and selection of general population students for group a. We have also proposed additional (optional) groups and conditions that the ETS research team will use to further strengthen the study. The outline of the design components follows:

High-level design. The design will consist of the three required groups of college freshmen taking the PARCC ELA Grade 11 Performance-Based Assessment (PBA) and End-of-Year Assessment (EOY) at agreed-upon university locations in fall 2014. We will then use these scores to predict first-year grade-point averages (FYGPA) and compare the predictions to those FYPGAs earned at the end of the 2014–2015 school year. Validity coefficients by group will provide evidence about the strengths of the predictive relationships for the three groups, and differential prediction analyses will indicate the accuracy of predictions for each group.

Assessment. The PARCC ELA Grade 11 PBA and EOY forms from the spring 2014 field test will be used for the study. For the field test, one ELA Grade 11 PBA form and one EOY form will include text-to-speech as an embedded accommodation. In addition, an alternate form containing the same items will not have TTS available. The field test forms will follow the proposed operational blueprint as best as possible given the limitations of the field test. The use of field-test forms will greatly facilitate completing the study in fall 2014, as the forms will already be published and human scoring of the tasks on these forms will have been completed.

Both PBA and EOY components will be administered in intact and separate sessions, mimicking the operational administration as closely as possible. With approximately three to four hours of testing for PBA and two to three hours for EOY, the total amount of time required per participant will be approximately six hours across two testing sessions. Although the PBA will be completed earlier in the school year than the EOY components for the operational PARCC, we plan to administer EOY first so that we obtain a complete set of EOY data.

Participants. We will select first-year college students in four PARCC states for participation in this study. In the first stage of sampling, we will recruit colleges and universities that are diverse with respect to selectivity and geographical location across PARCC states. We will try to include large public universities, private colleges, and community colleges so that the sample is representative of a distribution of "college ready" students. So that we achieve adequate sample sizes, we will target some institutions that have specialized programs for students with disabilities along with colleges and universities with typical services for students with disabilities. For example, we will target the University of Arizona's Strategic Alternative Learning Techniques (SALT) Center, which services 500 students per year and has a recruitment presence in four PARCC states (i.e., Arizona, New York, New Jersey, and Illinois), as a school that has specialized programs for students with disabilities.

We anticipate the need to recruit up to 20 colleges and universities to participate in this study. We will recruit schools by contacting both the Office of Institutional Research and the Disability Support Services office (or similar). We will contact each institution's Disability Support Services office first to determine the number of students supported by that office that are admitted or enrolled as freshmen. We will then narrow down these students to include only students who have indicated a need for the TTS accommodation in instruction and/or testing. We will invite those students to participate in the research study. We will select a sample proportionally similar to that sample in terms of broad undergraduate major (e.g., STEM, humanities, fine arts) from a potential pool of general population students who do not require accommodations. We will recruit the general education sample of students through the Office of Institutional Research.

The required and recommended groups are listed in Table 1, as follows.

Groups Required and Recommended for Inclusion in Study 3 Design				
Group	**TTS Provided**	**N**	**Session 1**	**Session 2**
Required Groups				
General Education	No	100	EOY	PBA
SWD who require TTS	No	100	EOY	PBA
SWD who require TTS	Yes	100	EOY	PBA
Recommended Groups				
SWD who do not require TTS	Yes	100	EOY	PBA
SWD who do not require TTS	No	100	EOY	PBA
General Education	Yes	100	EOY	PBA

As noted previously, we will administer EOY before PBA (the opposite of the operational ordering) in order to mitigate the loss of EOY data from second-session dropouts.

Table 1 includes six groups rather than the three required in the RFP. We include Group D, and its counterpart (Group E) for several reasons. First, opening the recruitment pool to any student with disabilities may encourage more accurate self-reporting by students with disabilities about whether they did use read aloud/TTS on an admissions test or require it in instruction. This would help to prevent students from reporting inaccurately in order to be included in the study (possibly for the monetary incentive). We could then ask students to self-disclose whether they do typically require the TTS accommodation, potentially providing some of the most accurate information that we will be able to obtain given that accommodated test sessions are no longer flagged. Second, inclusion of these groups would provide additional evidence for or against allowing TTS on a more widespread basis and validity evidence that provides a further comparison of predictive strength for various groups. We include a more thorough description of the proposed analysis approach, which also requires the inclusion of Group F, in the subsequent *Analyses* section.

Payments. Given the relatively small sample sizes available for this study and little incentive for schools or students to participate, we propose providing honoraria for both institutions and student participants. We will provide student participants with $25 per hour, or $150 for both testing sessions (PBA and EOY), with payment dependent on completion of both sessions.

We will compensate schools at the rate of $2,000 under the assumption that the Registrar's office will cooperate with the data collection coordinator in providing:

1. Email addresses for potential participants and social media postings informing students of the study details;
2. Course code mappings early in year 1 that will be useful in forming a standardized framework for comparison before data collection is completed;
3. Full-year course schedules for all participants at the close of the add/drop period in the Spring semester; and
4. Transcripts indicating end-of-year grades for the freshmen involved in the study in a timely fashion after completion of the spring semester.

Analyses. After we have collected the operational data and have received the transcripts from each school's Registrar's office, evaluation of data needed to analyze predictive validity may immediately begin, because all students will have taken the same form and raw scores may be used. We will complete the analyses that follow:

1. **Differential validity analysis:** We will obtain the correlations of the PARCC ELA Grade 11 field test (PBA, EOY, composite as recommended from psychometric analyses) with the criterion (FYGPA) by group, followed by typical statistical tests used to compare the strengths of the correlations (often transformed using Fisher's *r*-to-*z* test). The correlations, or validity coefficients, indicate our prediction that the PARCC score is of FYGPA in each group.
2. **Differential prediction analysis:** We will use the combined-group prediction equation to obtain residuals from that line for each group that indicate under- or over-prediction. This will provide evidence of whether the relationship of the PARCC as a predictor to the FYGPA criterion is similar or systematically different for the groups.
3. **Differential boost analysis (optional):** If the additional three recommended groups in Table 1 are approved, we can perform a differential boost analysis. This type of analysis requires either repeated-measures data (i.e., each student in each group takes the test under both accommodated and non-accommodated conditions) or independent-groups design with random assignment (as we recommend). The score boost is defined as each student's accommodated score minus his or her non-accommodated score. Score boost is then compared across groups. Whether the group requiring accommodations received a significantly greater score boost than the general population group provides further evidence of whether the accommodation is appropriate in this context.

Several parts of the study will be implemented in coordination with Pearson, who will be responsible for administering and scoring the field test forms used for the study.

Tasks requiring coordination will include: registration for testing sessions for participants recruited by ETS; administering the test form across two sessions at school test centers to the selected number of students (with targeted testing dates between September 15, 2014 and October 30, 2014); routing students based on login ID to one of two testing conditions (with or without TTS) and allowing students to use enlargement/magnification and other available supports as needed; providing item-level test data and information about whether or not the student used the TTS feature; and providing a list of completed test data on a weekly basis in order to facilitate payments.

In addition, we will work with Pearson to determine a mechanism for administering surveys to study participants. These surveys may be attached to the end of each test or possibly administered separately from the test administration system. We will target having the surveys TTS enabled for all conditions.

Human Subjects Review

Schools will be responsible for collecting parental permission for student participation. ETS will provide to the schools the contact letter and forms to be completed, which will outline the cognitive lab process. The protocols and instruments ETS uses will undergo fairness and sensitivity review to establish that content is appropriate.

ETS and Pearson have long been focused on establishing that research dealing with human subjects is of minimal risk to the participants and that proper procedures are undertaken to protect the confidentiality and welfare of those participants. The concern of ETS and its sponsoring organizations applies to all research conducted by ETS. ETS's institutional review board, the Committee for Prior Review of Research (CPRR), has a federal-wide assurance of compliance with the federal Department of Health and Human Services. The CPRR reviews research projects that involve human subjects for confidentiality, informed consent, and risk consistent with 45 CFR 46 and the Belmont Report. The CPRR consists of members from various parts of the organization (e.g., Research & Development, program areas, and the legal office) and has developed a process to implement the human subjects policies of ETS. The CPRR's area of concern is limited only to research projects that involve human subjects in some way (either through the use of existing data or through the collection of new data).

Since 2002, ETS researchers have been asked to sign the Research Agreement, which certifies that they have read the Prior Review of Research information about the human subjects review process. Researchers must acknowledge that they are responsible for completing and submitting documentation about the collection of data or the use of data that has already been collected for additional analysis and that they will not begin any data collection activities until they have received approval from the CPRR. The process uses an internal, password-protected SharePoint® site that allows for efficient distribution of information to and collection of documentation from researchers. Hundreds of IRB forms have been submitted through CPRR via secure SharePoint since 2008.

Technical Report

As with all ETS research studies, the final report will be reviewed by PARCC and will be reviewed through the ETS peer-reviewed process for technical review. This process establishes that all ETS research reports follow guidelines for fairness in studies that involve human subjects and have high levels of technical rigor and editorial quality. If approved by PARCC, the full report will also be submitted for publication as a PARCC technical report.

Quality Control and Use of Open-Source Software

ETS has a great deal of experience with data quality control measures before, during, and after data analysis and places a high degree of importance on verifying accuracy at each stage. Quality control procedures typically include checking sample sizes throughout the stages of processing (and reconciling discrepancies), performing reasonableness checks on scores and demographic indicators, double checking (checking by more than one analyst) of a subsample of responses, and cross-checking so that we obtain the same result from various angles or using differently attained data sets. The software to be used in analyses will be open-source or commercially available (e.g., R, SAS, SPSS).

Methods, processes, and procedures will be thoroughly reviewed internally by ETS researchers (including those with psychometric training and validity backgrounds) and data analysts. Further evaluation of these aspects will be achieved from presentation to and consultation with PARCC and its Technical Advisory Committee (TAC). Reports will be subject to a comprehensive technical review process, and ETS-approved reports will then be put through external review by the PARCC TAC and other external experts that may be suggested by PARCC prior to their final submission.

Study 4: Accessibility of New Items/Functionalities and Use of New Devices

The Partnership has identified a number of studies intended to verify that the assessment measures the Common Core State Standards validly and reliably and establishes a commitment to pursuing innovation and quality in the program's initial years. Study 4, dedicated to investigating the accessibility of new item interactions and functionalities and the use of new devices, is exemplary of the Partnership's desire to innovate with regards to fairness and validity.

A Two-Part Study: Overview

Pearson and ETS propose joining forces to pursue this study of new devices and new item interaction types through two series of cognitive laboratories: one focused on the accessibility/usability of new item interactions/functionalities and the other on new devices. While some overlap will exist between the device study and the new item interaction type study, they will be structured as separate endeavors that can nonetheless leverage and build on one another's findings.

The current range of PARCC-sanctioned devices will be used within the item type study, with greater emphasis on devices that are considered to be more vulnerable to device mode effects (e.g., tablets with their smaller screen size and lack of roll-over effects and cursor change-outs as user feedback mechanisms). The range of Partnership approved devices and our understanding of the potential for mode effects will most likely expand over time. This will be taken into account as we plan to conduct this study in years 2 and 3 of the contract. Students with disabilities, English learners, and students who are not a part of a special population will be included in the study. Students with disabilities participating in the item type study will use the devices that would be most typically used by them for the PARCC assessment and will take into consideration any peripherals or assistive technologies that would be used in conjunction with the primary device (e.g., wireless refreshable Braille display used with an iPad).

The device study will typically include a range of item types, including newer item types, but may be focused on devices that suggest a particular study design. For instance, if new on-screen typing or speech-to-text technologies seem promising and worthy of inclusion in a device study, a test focused on writing items may be indicated. Or, if tablets designed for interacting with tactile graphics are investigated, blind and low vision students may be the focus, with likely item types consisting of hot spot items and other interactions suitable for use with this device. With one study occurring in early fall and the latter in early spring, findings regarding a particular item type, potentially when accessed by students with certain disabilities, could lead to a theory regarding a particular vulnerability or opportunity that would be investigated within the device study. For instance, the early spring study might investigate a new device, peripheral, or assistive technology that could improve or resolve an access issue for certain item interactions discovered in the fall study. Or, certain new item types studied in the fall may be highly promising but require an amount of screen space that contraindicates the use of smaller device sizes, despite possible popularity in schools and the consumer market. The spring device study could then try to identify where the "too small" line is drawn, using knowledge from the fall regarding item types requiring ample screen space.

Division of Work Responsibilities

This effort will be a highly collaborative in order to take advantage of complementary research expertise from Pearson and ETS. Pearson and ETS will work together on the study design, divide data collection responsibilities, and coordinate on findings, conclusions, and recommendations. Pearson will take responsibility for the publishing of forms and the purchase or renting of hardware. ETS will take responsibility for garnering approval from the Institutional Review Board. Program management oversight and support will be supplied by Pearson. Content expertise for new item interactions being studied will be based on subject and grade level, with ETS content staff involved for the subjects and grade levels for which they are doing item development and Pearson content staff involved for the remainder. Within this collaboration, each party will review the other one's work in order to build basic quality control measures into each stage of the research.

Test Delivery System

The new item interaction/functionality study will use conducted using TestNav.
Full security will be maintained for PARCC items used in the study, with sensitivity to item exposure. The form for the study may be made available through a secure practice test format.

Year Two Activities

Pearson/ETS proposes that this program of research will be a part of an integrated and rationalized approach to the development of new item interaction types that becomes solidified in the first year of this contract. As part of an accessibility-first approach, we propose that an accessibility plan will be drafted for each proposed item type. This plan will include considerations for touch-screens; interaction with text-to-speech as well as other system-based accessibility tools; any necessary interaction with assistive technology; and any special populations for whom this item type would not be the most appropriate mechanism for measuring certain constructs. Prior to technical implementation of a new item interaction type, the accessibility plan for an item type will be presented to the Partnership along with the proposed item data model and scoring strategy, since modifications or additions to the PARCC item data model are subject to Partnership approval. ETS and Pearson accessibility experts would be involved within this endeavor, with the hypotheses around the methods for achieving accessibility with a given item type becoming inputs to the research design for Study 4.

We propose that the Partnership consider an additional component of this study that would be an investigation of accessibility in custom interactions and portable custom interactions (PCIs) using the Accessible Portable Item Protocol (APIP). Currently, the method for APIP tagging does not apply smoothly to PCIs due to PCI's (and some custom interactions') use of JSON and CDATA structures rather than more traditional HTML structures used for attaching APIP tags and exposing text-based content to text-to-speech technologies. Additionally, PCIs programmed independently of the test delivery system may not interact as intended with system-based accessibility measures. While these challenges are not insurmountable, any possible barriers to supporting both interoperability and accessibility in new item type interactions will need to be resolved early within the PARCC operational timeline.

Additionally, ETS and Pearson would like to propose possible methods for understanding optimal numbers of different item interactions encompassed within a single ELA or math test at a given grade level. PARCC research conducted to date suggests that younger grade levels may have limited ability to respond agilely to a wide range of differing item interactions in a single test. However, with an opportunity to learn and practice item type interactions, this tolerance for experiencing many different item interactions in a single test may expand. Gradual introduction of new item interaction types over multiple years may prove to be an acceptable way to expand the number of item types. The expansion of item interactions over the subsequent years could then take place with some sense of appropriate limits and necessary opportunity-to-learn measures.

General Design

Both studies will use a similar cognitive laboratory approach combining two methods: (1) a usability protocol that relies heavily on observation and highly specific coding of success criteria, item type learnability, and usability metrics and (2) cognitive interviews intended to supplement the limited concurrent verbalization (or "think aloud") that typically occurs when students are absorbed in problem-solving and concentrating on content issues. The cognitive interviews will provide a retrospective view on student actions and ask students to reflect on certain aspects of the tasks. Where problem-solving steps are not immediately observable, students' approach to item understanding, problem processing, and response creation will be revisited in the cognitive interview. Preparation for the study will have involved coordination with content staff to understand common misconceptions, expected steps, and differences between novice and expert approaches associated with a construct that is intended to be measured by a new item type. The cognitive interview will provide an opportunity to analyze a student's approach to an item in relation to this information about content knowledge and process skills.

Both ETS and Pearson have had success with this model within past studies and have independently arrived at similar refinements to the approach, particularly when working with English learners and students with disabilities. This methodology is very cost-effective, and the resulting recommendations tend to be highly actionable, whether in terms of content creation guidelines or software redesign, and available in a short timeframe, thereby maximizing the potential for results to be used within the current year's efforts.

Item Types: Research Questions

The questions to be targeted in the item interaction and new functionalities include the following:

- Do the new items functionalities facilitate or inhibit students' ability to demonstrate their knowledge and skills in relation to the construct being measured?
- How do student paths through new item interactions compare to hypotheses around how an item type may differentiate between students with and without the knowledge and abilities that are intended to be measured?
- What type of effort is required for students to gain full understanding of how to use item functionality?
- Are any usability issues encountered?
- If possible to effectively discern, is some portion of student cognitive processing dedicated to understanding and working with item type functionality such that it appears to distract from full engagement with the construct?
- In the case that an existing item type is being brought to this study with new accessibility features in place, are those accessibility features successful in providing access to special populations?

- Are the features and measures described in the item interaction accessibility plan in place and successful in providing access?
- Do particular usability, access, or construct processing issues arise for certain populations when interacting with this item type?
- Is there any indication that the way an item type measures a construct is an inappropriate match to how some students with disabilities would learn that construct and be assessed in an instructional setting?
- Are there incompatibilities or unintended effects in how item type functionality works with assistive technologies?
- In the case that non-item-type-specific embedded supports are built into the test delivery system, do those supports work as intended within new item interaction types?

Item Types: Sampling Plan

Pearson/ETS proposes that 20 students from each of three grade levels be recruited within two states for the purposes of this study.

Two states	X 2
Three grade levels	X 3
At each grade level within each state	20
General population students	4
Two students from 4 disability categories	8
Two students from each of 3 English proficiency levels	6
Two students previously EL recently reclassified	2
TOTAL	120 students

ETS would perform data collection in one state, preferably a northeastern state. Pearson would perform data collection in the other chosen state. The Partnership and Pearson/ETS will solicit interest from state departments of education, providing a description of the research and the type of required commitment for the study. Using contact names provided at the more local level, Pearson and ETS will follow up to arrange the logistical details.

To encourage participation, Pearson/ETS recommends providing honoraria as follows:

- $500 honorarium per school for up to three students plus $100 for each additional student
- $20 honorarium to each high school student participant

Schools will assist in the distribution and collection of a permission form that must be signed by parents in order for a student to participate. Schools will be asked to provide a suitable location to accommodate up to four researchers. School devices could be used, or Pearson/ETS can bring devices, assuming that it will be possible to access the school's wireless network.

Students from the general population, English learners (ELs), and students with disabilities (SWDs) will be included in the study. Cognitive labs with general population studies will permit an understanding of any usability issues and provide a baseline for comparing SWDs' and ELs' ability to access those item types in an equivalent fashion.

Item Types: Methodology

A form will be constructed for each grade level that will consist of new item interactions and, if relevant, previously existing interactions that have had additional accessibility functionality added since last being studied in a cognitive laboratory format. The form will combine ELA and math items, assuming that new item types exist in both subject areas. New item types will only be tested at the grade levels for which they have been deemed appropriate. Attempts will be made to work within the prescribed study format in order to study new item types at the lower end of the range for which it is appropriate. While a new item type may be studied at multiple grade levels, some efficiencies may be achieved by understanding that if an item type performs well with students at a certain grade level, it is likely to perform well with students above that grade level too.

When possible, two items of a given item type will be included in order to (1) understand how the item type functions when used with different content and (2) identify any differences between the students' interaction with the first and second item, assuming that the student is learning the functionality in the first item and making use of that knowledge in the second item. The greatest challenge in assembling forms will be limiting the number of passages in ELA, since the form should be able to be completed within a class period or no longer than 75 minutes, including time for a post-hoc cognitive interview.

Students will receive a form designed for the grade level that they completed in the spring rather than their current grade level. Before the beginning the forms, the students will be introduced to the researcher and given an overview of the study and what they should expect within the session. Researchers will observe while the student is working through the items. Students are welcome to reflect on any part of the interface that they find notable as they work on the items, but the facilitator will allow students to work on items undisturbed.

For each item, the protocol will be used as a framework for observations in addition to a freeform area where the researcher can make observations about behaviors that may not have been predicted and included in the protocol.

New Devices: Research Questions

The research questions for the device study will be largely driven by the nature of the device, peripheral, or enhanced capability. A survey of new technologies will initiate the work, with Pearson bringing to the Partnership a short list of technologies, each with a description and associated research questions. Input from the Partnership will be used to narrow the list to two technologies. Criteria for selecting the technology or device might include one or more of the following:

- Could this device make accessible item types that have not been accessible to students with certain disabilities in the past?
- Does this device exist outside of PARCC guidelines but offer some promise in terms of cost and/or availability in the classroom?
- Does this device or peripheral offer a new way of interacting with content that should be considered by the Partnership?
- Does this capability offer some potential advantage that may need to be investigated in order to understand comparability?
- Does this device or capability resolve some issue that has been observed in the past (e.g., inaccurate or slow typing, reduced user feedback, input imprecision)?

New Devices: Sampling Plan and Methodology

The devices study is anticipated to be smaller in scale, occurring in two states with six students at each of two grade levels for each of the two chosen technologies or devices, for a total of 48 students. The exact study design in will be highly dependent on the devices chosen.

Device	Students Recruited	Form Specifics
INtact Sketchpad	Blind students within 2 grade ranges	Items suitable as interactive tactile
Research question(s): Can such devices make accessible to blind students technology-enhanced item types such as hot spots, when delivered as interactive tactile graphics?		
On-screen keyboards	Students who have become adept at using new non-QWERTY keypads	Open response items and a small range of other item types
Research question(s): Are students successfully using some of the newer non-QWERTY on-screen keyboards? Do these keyboards offer certain advantages such as limited occlusion of the screen?		
New smaller, low-cost device (minis)	Six students from 2 grade levels	Emphasis on item types expected to be more challenging on a smaller screen

Device	Students Recruited	Form Specifics
Research question(s): Does this device present usability issues that would hinder or restrict students' ability to interact with and respond to assessment items and undermine comparability?		
Haptic feedback	Six students from 2 grade levels including low-vision students	Item types most likely to benefit from haptic feedback
Research question(s): Does haptic feedback increase the usability of items on a touch-screen? Would this benefit low-vision students in particular?		
Speech-to-text	Motor disabled, ELs, non-spec. pop.	Writing and fill-in items
Research question(s): Do writing items used with speech-to-text measure writing constructs? What populations may benefit from this capability being built into the test delivery system? In what situations is this technology not a suitable replacement for assistive technology?		
Convertible laptop-tablets	Six students from 2 grade levels	Item types to benefit from additional precision from touchpad mouse
Research question(s): How easy is it for students to transition between tablet and laptop configurations? Are there specific item types where students prefer one configuration over another?		

Study Deliverables

The deliverables that will be provided to the Partnership in association with these two studies include the following:

- For the device study, a short list of technologies and devices to consider studying in the current year, including information about each technology/device, research questions, and implications for sampling
- A draft of each study plan, including timelines, study design, sampling specifications, and data collection methods
- A final study plan
- Access to each form being used within a study
- A draft general protocol and an item-specific protocol to be used within the study
- A final general protocol and an item-specific protocol to be used within the study
- A draft study report
- A final study report, including theoretical framework, design rationale, executive summary, and specific recommendations of action
- An action plan based on recommendations in the final study report, indicating responsibilities of each party involved along with a timeline for each action
- Update reports on that action plan

Draft Milestone Schedule

Task	New Item Types	New Devices
List of possible technologies to study	N/A	June
Study Plan Draft	Early May	Early September
Final Study Plan	Early June	Early October
Item Selection	June	October
Forms Ready for Review	Early July	Mid-November
Draft Protocols	Early July	Mid-November
Final Protocols	Early August	Early January
Recruitment Materials Finalized	Early August	Early January
Recruitment	August	January
Local Logistics Negotiated	September	February
Study	October	March
Draft Report	November	April
Final Report	December	May
Action Plan	January	June
Updates on Action Plan	As needed	As needed

Dates proposed were based on original plan to conduct study in year 1. These dates can be adjusted to reflect administering the study in years 2-3 only. Final dates for draft study plan and subsequent deliverables to be negotiated once contract start date determined.

Risks

A number of risks and mitigation strategies can be identified:

- Recruitment of students with disabilities can sometimes be difficult. Recruiting in two or more different states will be helpful in that if some disability gap exists in one state's recruitment efforts, that gap may be filled in the other state.
- Some students may , have only had a partial opportunity to learn CCSS. While content knowledge gaps can make it more difficult to understand functionality when functionality is highly content-specific, general usability data can still be obtained.
- The timing for revision and approval of items may not align well to the item type study when limited items using the new item types are available. An expedited review and/or items not moving through the entirety of the review process may need to be considered.
- The test delivery system may not be fully compatible with new devices such that an alternate way to examine the utility of the device may need to be considered.

Study 5: International Benchmarking Study

It is important to examine how PARCC students' performance compares with that of their peers in top-performing nations across the world. Information from the international comparisons will help business leaders, governors, parents, educators and the public at large

better understand the performance and progress of the education system in each state. In addition, international benchmarking studies may provide valuable data to evaluate the rigor of future PARCC performance standards (i.e., cut scores for performance levels).
Such comparisons will be conducted by establishing statistical linkages between PARCC assessments and TIMSS and PISA 2015 assessments. In the pages that follow, ETS has outlined two methods that can be considered to achieve this goal.

1. **Embedded administration of TIMSS and PISA item blocks in the PARCC Test**. This would entail administering TIMSS or PISA item blocks in the matrix selection of the PARCC operational test forms to a sample of students. The PARCC score scale can then be linked to that of TIMSS and PISA by a variety of methods. However, all of these methods make important assumptions, and the validity, generalizability and defensibility of results will depend on the degree to which those assumptions hold.
2. **Pseudo-equivalent groups**. No TIMSS or PISA items would be administered by PARCC. Instead, the PARCC population is demographically weighted to be equivalent to the TIMSS and PISA national samples. PARCC scores are then linked to TIMSS/PISA by equivalent groups methods. This method avoids all mode, IRT model, test timing, and time of year of administration issues because both tests are administered in the appropriate mode at their traditional time of year. However, this approach too makes strong assumptions, and the validity of the results depends on the degree said assumptions hold.

There are several types of information that can be gleaned from international comparisons. These range from information about performance on common items; concordance information that allows us to map a cut score from one test onto the other test's scale; and projection data that gives a prediction of how one would score on one measure given a certain level of performance on the other. In addition, there are different ways to obtain each type of information, and these diverse methods can rely more or less heavily on assumptions about factors like delivery mode effects, context effects, the impact of different testing windows and years, and the implications of different content frameworks. There are also ways in which TIMSS and PISA might be used judgmentally during standard setting that do not rely on empirical linking data per se. Such possibilities are described in our response to section V.E.5.A. (Benchmark Study to Inform PARCC Middle and High School Performance Standards). In any condition, it is important to note that these types of information are only available under certain conditions and that these conditions cannot always be manipulated or influenced. Each of these options requires choices in prioritization of the information most valued.

Practical considerations often require linking activities to be conducted under less than ideal circumstances. The TIMSS and PISA samples will not contain substantial numbers of students from each PARCC state. The TIMSS Math and PISA Literacy content and skills frameworks are also quite different from the PARCC math and ELA frameworks. Finally, TIMSS and PISA data both come from paper-based administrations; while PARCC assessments will be on both computer and paper (PISA plans computer administrations for 2015). Given the goal of predominantly computer-based administration, we recommend use of the PARCC online assessments for this study. We will work with PARCC and its TAC to adapt the proposed study design to best meet the goals of the PARCC assessment system.

Different Approaches to Providing Information for International Benchmarking

The various sorts and sources of information useful to conducting and evaluating the linking between PARCC and international assessments are reviewed below. The focus here is on information that will allow us to determine the extent to which PARCC proficiency standards are aligned with those established with TIMSS and PISA. Although this determination could be made entirely by judgmental methods, an empirical link between the PARCC score scale and that of TIMSS and PISA is certainly preferable.

Under ideal circumstances, the design for establishing score links between PARCC and the relevant TIMSS and PISA measures would include the following elements:

- A paper administration of TIMSS and PISA
- A computer administration of TIMSS and PISA
- Computer administration of the PARCC test
- Some sample of students getting both PARCC sections and TIMSS/PISA blocks on computer (to allow for concurrent calibration)

This design would allow us to examine the possible computer versus paper differences on PISA as well as give us equivalent samples on TIMSS/PISA and PARCC.

In 2015 we will have a naturally-occurring sample of students and schools taking both TIMSS or PISA and PARCC. This may afford the opportunity to validate or enhance the data used to support the 2015 standard setting. It is also worth noting that having some of the PARCC schools take TIMSS or PISA on paper would allow use of methodology like those employed by McLaughlin & Bandeira de Mello, (2002) and Braun and Qian (2007) to map state assessment cut points onto the TIMSS scale.

There are essentially two types of approaches available to us in the 2015 time frame: embedded items and equivalent populations. Both have advantages and potential drawbacks, with neither being ideal.

(1) Embedded Items

A variety of embedded-item designs are possible, differing both in the number of embedded-items employed and in the way they are distributed across students. Preference for a given design depends on the nature of the tests being linked and the circumstances under which they are administered. Some tests are more easily linked than others. Tests that measure the same construct, use the same item types, and are administered under similar conditions to similar populations are more easily linked than are tests that differ in some (or all!) of these particulars. Embedded-item designs can also be characterized as stronger or weaker, with stronger designs generally using larger common-item sets and administering more of that set to each student. Stronger linking designs make more tenable assumptions and are more robust to violations of those assumptions.

They are therefore more likely to yield results that generalize across different samples of items and students. The goal is to appropriately match the strength of the linking design to the difficulty of the linking problem. .

Establishing links between PARCC and the TIMSS and PISA score scales will be, by almost any measure, challenging. The PARCC ELA test, which jointly measures reading and writing, will be linked to TIMSS and PISA tests that measure reading exclusively. Although the math constructs are better aligned, the nature of the items used varies substantially across programs. Both TIMSS and PISA are (currently) paper-based tests while PARCC is mainly administered on computer. The tests are also delivered at different points in the school year. TIMSS is traditionally administered in spring, while PISA is delivered in October and November. In contrast, PARCC will generally test between March and June. Finally, PISA targets its population by age rather than grade level. Although no embedded-items design is likely to cope completely with all of the above challenges, a stronger design is certainly more likely to produce a satisfactory outcome. .

The strongest linking design would employ the largest feasible common-item set and administer the largest possible chunks of it to each study participant. The ideal would be to embed the equivalent of an entire TIMSS or PISA administration, a total of several hundred items. Although each student would be administered only a fraction of the TIMSS or PISA items that are embedded, enough blocks of items would be spiraled to give solid coverage of the TIMSS and PISA frameworks. It would also be ideal if each participant's test session contained roughly equal numbers of PARCC and TIMSS or PISA items. Administering large blocks of items of each sort better allows that each block adequately samples the constructs measured. The administration order of the PARCC and TIMSS/PISA sections would ideally be counterbalanced, with either appearing in the first position as often as in the second.

There are formidable practicality and cost considerations associated with embedded TIMSS and PISA items into the PARCC assessments. Delivery considerations will likely limit the number of linking items that are administered to each participant, spreading the common items more thinly across students. The need to restrict the linking sets weakens the design and risks impacting the utility of the data collected. It is likely that the impact of design weaknesses will fall harder on the ELA linkage than on math, where the constructs are better aligned.

The source administration of available TIMSS/PISA items will also have implications for the analyses and the inferences that can be drawn. Although TIMSS and PISA will be administered in 2015, it is not clear that these items will be available in time for this study.

A risk inherent in any variation of the embedded-item approach lies with the assumptions it requires. Since we will conduct the comparability study on the computer-based administration which is the targeted administration mode of PARCC, chief among these is the assumption that the (currently) paper-based TIMSS and PISA items can be successfully converted to computerized administration.

To the extent that such conversion systematically changes item performance, they will not act as anchors between the PARCC and TIMSS/PISA score scales. A secondary concern is that the TIMSS and PISA items were last administered in 2013 and 2012, respectively. To the extent that 2015 students will not respond to those items like past students did, the linkage might again be negatively impacted. Although these assumptions will be checked in 2015 when TIMSS conducts a mode effect study and re-estimates national and state proficiencies, this will obviously occur after the PARCC standards are set and implemented.

(2) Pseudo-Equivalent Groups

Given the challenges inherent in the embedded-items approach, statistical adjustment of observed samples is a possible alternative. The idea here is to make use of standard methods that have been successfully used for decades to produce what might be termed pseudo- or synthetically-equivalent groups. These methods are often termed “post-stratification” (Cochran, 1977). “Propensity-score matching” is a related methodology that is being applied to educational studies with increasing popularity (Rosenbaum & Rubin, 1983). In our situation, the process would work as follows:

1) TIMSS/PISA score distributions (for reading and math) would be taken either in their full, nationally-representative form or determined for the specific collection of states that participate in the PARCC test. The relevant demographic characteristics of the sample would be tabled. These would include both student characteristics (e.g., ethnicity, gender, SES) as well as district or school characteristics (e.g., urban, suburban, rural). For PISA, the same approach can be applied at the regional level.
2) The corresponding demographic characteristics of the PARCC standard-setting sample would also be tabled, along with the (projected) PARCC scaled-score distribution.
3) Post-stratification methods would be employed to map both the TIMSS and PARCC samples to a common, synthetic population on which demographic characteristics are most equivalent.
4) The same weights that take both the TIMSS and PARCC samples to the synthetic population would be applied to the two score distributions (with appropriate smoothing, as necessary), emulating the outcome of administering both tests to common or equivalent groups.
5) At this point, standard equivalent-groups equating methodology would be brought to bear, with score values on one test matched to values at the same percentile rank on the other.

This approach has notable advantages over embedded-item linking, including:

- Tests are administered (or have been administered during their regular schedule) in their appropriate mode (computer for PARCC).
- Tests are administered at their appropriate points in the school year (e.g. March–June for PARCC, and October–November for PISA).
- Item context, position, and response modeling issues are eliminated.
- Issues surrounding the size, representativeness and security of common-item anchor sets are also rendered moot.
- Analyses are based on standard, time-tested methodologies.

The main drawback of this design rests in the assumption that the post-stratification adjustments in fact produce something akin to equivalent groups. This depends both on the availability of relevant demographic characteristics and on the assumption that these characteristics relate to student performance in the same general ways across testing programs. Furthermore, if the differences between the observed samples are large, post-stratification is unlikely to produce plausibly equivalent groups. The design also shares a weakness with the embedded-items approach in assuming that overall levels of student performance remains essentially unchanged since the last PISA and TIMSS administrations.

in considering the relative merits of the two approaches, the embedded-item approach offers the promise of strong linkages, but only if administration mode, item context, test timing, and time of year impact item performance in minor ways. The pseudo-equivalent groups approach is, in contrast, largely immune to these factors, and although it is subject to other statistical weaknesses, as noted above, it is eminently feasible and has provided meaningful results in previous studies. Thus, we recommend the pseudo-equivalent groups approach, recognizing that PARCC and its TAC will likely revisit the two possible options at an early point in the project.

Coordination with Other Partners

If the option to embed TIMSS or PISA items into matrix sections of the PARCC assessments is pursued, ETS will need to work with PARCC and Pearson to determine the impact on operational testing. For either of the study options discussed, PARCC will need to secure the cooperation of TIMSS and PISA officials and their contractors for obtaining item permissions and associated scoring rules/rubrics.

Sampling and Missing Data Considerations

One of the sampling issues that will arise with embedding TIMSS or PISA items into the PARCC assessments is that the sampling of items will be very challenging. As previously discussed, it may be difficult to sufficiently represent these measures in a limited number of matrix sampling sets.

Quality Control Procedures

The processes and procedures will be reviewed both internally by other staff members from the Pearson and ETS teams and externally by the PARCC TAC. Quantitative data and analysis results (e.g., item-level statistics, reliability and validity estimates, etc.) will be verified by multiple researchers. Reports will undergo a multi-stage iterative review process including internal review among team members, external review, PARCC review, and final submission.

V.C.6. Technical Documentation for Assessment Administration

Requirement

Response Requirements for Section V.C.6.

a) Descriptions of the approach and procedures to complete all the responsibilities/tasks specified in Section V.C.6.

Deliverables for Section V.C.6.

a) Draft and final technical manual that provides all of the information in the outline below, as well as any other analyses identified by the Contractor and deemed appropriate for the report by Partnership representatives
   i. The Contractor shall complete the draft manual within 3 months of the end of the test administration or a date mutually-agreed upon by the Partnership and the Contractor.
   ii. PARCC will provide feedback on the draft and the Contractor shall complete the final document within 1 month of the PARCC feedback or a date mutually-agreed upon by the Partnership and the Contractor

Response

The technical manual will serve as the central repository for the technical documentation related to the assessment system. It will inform readers of the rationale and framework for the assessment system, showcase the foundational psychometric and research work done to inform its design and development, present the validation argument underlying the system, and provide operational criteria and data that show how the assessments are functioning.

As the psychometric lead for the project, ETS will take responsibility for the design and development of the technical manual. Pearson will provide inputs to several of the technical manual sections. Measured Progress, in its role as independent auditor, also may provide evaluation of or input into the technical manual as it is developed.

While of the elements outlined in the RFP will be included in the technical manual, ETS proposes a slight revision to the order of the elements as well inclusion of some additional information. However, ETS recognizes that the final design of the manual will depend upon the input of PARCC and its technical advisors, and may evolve over time as new data and information become available to the program..

We propose that the technical manual comprise four main sections, each containing multiple subsections around a common theme, as follows:

- **Section 1** will provide an overview of the design, validity framework, and claims of the assessment system as well as summaries of studies conducted during the initial design phase.
- **Section 2** will focus on operational and development issues , describe design processes (such as Evidence Center Design and Universal Design), and provide information relevant to item development, item banking, test form construction, test administration, scoring, and reporting.
- **Section 3** will present technical and psychometric information, including topics ranging from item analyses, DIF, equating and scaling, and standard setting methods. We also propose that operational performance data for the overall group be included in this Section.
- **Section 4** will address validation work, including summaries of completed studies and other evidence in support of the validity argument. Appendices will include information on test blueprints; present sample score reports; provide raw-to-scaled score tables; present statistics not included in Section 3; and provide data relevant to subgroup performance and other statistics.

In addition, we propose using an ETS-developed research framework to track, monitor, and propose relevant research. The framework contains 11 categories important to supporting the validity, quality, and fairness of an assessment. The framework will be included in the Appendices of the manual and will be used to document research underway, reference research and relevant program documentation that is completed, identify future research that is needed, and propose priorities for future research by year.

Each section will contain a preface summarizing the information included in the section. In addition, overviews of particular methodology or criteria will be provided as appropriate (for example, overviews of DIF, standard setting methods, the ECD approach, and item analyses). It is anticipated that each section will largely be independent from the others so that updates, when needed, will be made easily to individual sections with little disruption to other sections. This design will also allow readers to search and review particular sections online. The newest version of the professional standards for educational and psychological testing (expected to be available in prepublication form in December 2013) will be referred to, as appropriate, throughout the manual. In addition to the professional standards, other standards may be consulted and referenced. For example, *ETS Standards for Quality and Fairness,* which provides an operational interpretation of the professional standards, and *Operational Best Practices for Statewide Large-Scale Assessment Programs* from CCSSO and the Association of Test Publishers which contains descriptions of best practices for operating testing programs, may also be referenced as appropriate.

The manual will be authored as an e-book that will enable readers to easily search for information. It will be developed using the EPUB 3 standard that will allow converting the manual to almost any format needed (such as a PDF document or website). EPUB 3 includes accessibility from the ground up and as such, is the standard to use to provide accessibility in a digitally published document. The EPUB 3 standard was developed by the International Digital Publishing Forum (see http://idpf.org/). ETS representatives participate in working groups sponsored by standards organizations that focus on information accessibility (such as IDPF, W3C, and IMS). Therefore, we propose creating the technical manual as an e-book, following procedures that provide that the document meets accessibility guidelines and standards (e.g., http://www.idpf.org/accessibility/guidelines/; http://www.w3.org/TR/WCAG20/).

The proposed outline for the manual is detailed below.

Section 1: Development, Design, and Framework of the Assessment System

This section will focus on the rationale and content of the assessment system. Information on pilots, field trials, and other data collection efforts will be presented; the claims made by the assessment results described; summaries of research studies that are completed or in progress provided; and the relationship of the content of the assessment system to the Common Core State Standards described. In addition, future studies needed to support the underlying validity argument will be identified using the research framework described above. The following subsections will be included:

I. Overview of the Assessment System

- A. Purposes of the system
- B. Uses of the assessment information
    1. Inferential target(s)—school, student
    2. Uses of assessment results (including accountability)—state, school, and student

II. Content of Assessment System

- A. Brief overview of the Common Core State Standards
- B. Translating the Common Core State Standards into assessment specifications

III. Building the Validity Framework and Argument

- A. PARCC's validity orientation and framework
    1. Understanding the claims of the assessment system
    2. Connections among the content, learning models, and assessments
    3. Studies examining validity questions for PARCC
        - a. Completed

b. Near-term

c. Future

IV. Initial Development Efforts

A. Item Development

1. Item prototypes

2. Performance Based Assessments

a. Design of rubrics

b. Summaries of pilot studies and field tests

3. Technology Enhanced Items

a. Design of rubrics

b. Summaries of pilot studies and field tests

4. Other student-task interaction studies

5. Interaction of reading and writing

6. Links to ECD

7. Cognitive complexity

B. AI Scoring

1. Summaries of feasibility studies

Section 2: Test Development and Operational Specifications

Section 2 will focus on aspects relevant to the implementation and maintenance of the ongoing assessment program. The section will include an overview of the different test development approaches used in creating the assessments; specifications for items, item banking, and form construction; and descriptions of the processes for administering, scoring, and reporting assessment results. The following subsections will be included:

I. Evidence Centered Design and the Assessment System

A. Description of ECD used with the Assessment System

B. Models of expected domain proficiency

C. Development of ECD claims (also related to III.A.1.)

1. Description of major claims

2. Description of subordinate claims

D. Performance Level Descriptors (PLDs)

II. Accessibility and Accommodations

A. Universal Design for Learning

B. Accessibility studies

C. PARCC accommodations manual

D. Supporting claims regarding effective and appropriate accommodations

E. Test delivery and embedded accommodations

III. Item Specifications

A. Math item types (I, II, and III)

B. Reading (text complexity; informational and literary genre)

C. Writing (argument and textual writing)

IV. Item Banking and Meta-Tagging

A. Accessibility, Portability Item Protocol (APIP) specifications and item information

B. Question and Test Interoperability (QTI)

C. Interaction with Personal Needs Profile (PNP)

V. Form Construction

A. Assessment specifications

B. Test blueprints

C. Number of forms developed

D. Form design

1. Number of operational items
2. Use of embedded field test
3. Matrix linking items

E. Paper based forms

1. Creating comparable design
2. Use of "paper clones"

VI. Administration and Training

A. Administration procedures and guidelines

B. Professional development and training programs

C. Monitoring and quality control of administration procedures

D. Administration irregularities

1. Definition of administration irregularity
2. Procedure for handling exceptional incidents

E. Computer Based Testing

1. Description of process and procedures
2. Test sites certification
3. Handling software/ hardware failures

VII. Scoring

A. Scoring rules and criteria

B. Scoring process

1. Sections adjusted for automated versus human scoring
2. Use of anchor/training papers
3. Selection and training of scorers

C. Scoring quality control

1. Monitoring of scorers
2. Scoring accuracy
3. Scoring consistency
4. Scorer drift protocols and analyses
   a. Trend scoring (as applicable)
   b. AI scoring as quality control (as applicable)

VIII. Reporting

A. Report Design Process (ad-hoc reporting group; design Contractor)

1. User experience research (personas, interviews, focus groups, etc.)

B. Types of PARCC reports and purpose of each report

C. Types of Scores Reported

Section 3: Technical Criteria

Psychometric and technical information will be housed in Section 3. This section will include brief overviews to various psychometric approaches (e.g., item analysis, standard errors, vertical equating, DIF, standard setting, etc.) in appropriate subsections; provide information as to the approaches used and the outcomes of the analyses; and present basic performance and other statistics for the overall group. The following subsections will be included:

I. Alignment

A. Conceptualizing cognitive complexity

B. Traditional and alternative conceptions of alignment

II. Item Level Analyses

A. Traditional item analyses (e.g., difficulty and discrimination)

B. Examining DIF and item bias

C. Summary test characteristics for Overall Group (e.g., p-values, point-biserials)

III. Characterizing errors associated with test scores

A. Uses of test scores and implications for consideration of error

B. Levels of analysis

C. Decision consistency and accuracy

D. Generalizability analyses

E. Overall and conditional standard errors for Overall Group

F. Traditional reliability analyses for all reported scores for Overall Group

IV. Performance Statistics (for Overall Group)

A. By scale

B. By Performance Level

V. Calibration and Scaling

A. Calibration models and methods

B. Considerations in choosing a reporting scale for PARCC

1. PARCC score scales and characteristics

C. Interpretative quality of the scale

D. Vertical scaling study results

VI. Equating

A. Scale stability (comparability of scores within year)

1. Equating methods

2. Comparability across various administration forms

a. Paper to computer

b. Comparability across various devices

B. Comparability of scores across years

1. Equating methods

2. Evaluation of equating results

C. Vertical linking studies

VII. Standard Setting

A. Research to support standard setting

1. Benchmark study to inform PARCC middle and high school performance standards

2. Performance of post-secondary student study 4/25/2014 1/24/2015

3. Postsecondary educators judgment study to inform cut scores in PARCC high schools assessments

4. Field trial of standard setting study

B. Standard setting methodology

1. Rationale for method used

2. Performance descriptors

3. Panelists

4. Protocol

C. Standard setting results

1. Unadjusted results

2. Adjusted/smoothed results

3. Policy decisions

4. Coherence across grade levels

5. Coherence across subject areas

6. Policy decisions

7. Validity of standards and cutscores

Section 4: Validation

Section 4 will present empirical evidence for the validation of the assessment. While the information presented in this section also relates to information provided in Section I, Part III, the intent of Section 4 is to provide the connections between the claims, development processes, and research studies that support the validity argument underlying the system. The following subsections will be included:

I. Empirical Evidence from Research Studies

A. Content-related (including both development processes and research studies)

B. Internal structure

C. Response processes

D. Relationship to other variables

E. Consequential

II. The Validity Evaluation

A. Revisiting the validity evaluation questions and claims

B. Logical/theoretical relationships among the content, students, learning, and assessment—revisiting the assessment triangle

C. Synthesizing and weighing the various sources of evidence

Appendices

A. Test blueprints

B. Administration procedures

C. Sample reports

D. Performance statistics by subgroup (e.g. means, standard deviations, percent in PL)

E. Summary test characteristics by subgroup (e.g. standard error of measurement, average p-values, average point-biserials, etc.)

F. Detailed IRT and classical item level information (e.g., IRT calibrations, error, p-values, point-biserials)

G. Raw to Scale Score Conversion Tables with associated conditional standard errors of measurement

H. Reliability coefficients by subgroup

I. Test characteristic curves

J. Test information functions

K. Research framework for ensuring validity, quality, and fairness (see example below)

Category of Research	C = Completed study O = Ongoing study P = Proposed study	Year of study (or proposed year)
1. Validity Evidence: Provide evidence that supports the intended inferences and actions based on the reported results for a testing program		
2. Fairness and Accessibility: Providing quality and validity for all test-takers, including those with disabilities or English-language learners		
3. Support of Ongoing Program Change: Conducting foundational research, analytical trials, and item development issues that support test revisions or maintenance		
4. Scores and Scales: Evaluating scale concerns, the maintenance of an existing scale, and the soundness of different scores		
5. Security: Examine security issues that impact the underlying construct of a test		
6. Scoring and Technology: Applying technology in the support of test scoring		

Category of Research	C = Completed study O = Ongoing study P = Proposed study	Year of study (or proposed year)
7. Test Preparation: Evaluate the impact of test preparation on test performance		
8. Candidates and Populations: Who is taking the test and performance differences		
9. Psychometric Properties: Examining psychometric characteristics at the item and test level that affect quality and validity		
10. Score Interpretation: Evaluating score inferences and their use		
11. Policy Issues: Application of information from the test in decision-making		

V.C.7. Quality Control

Requirement

Response Requirements for Section V.C.7.

a) Description of the approach and procedures to complete all the responsibilities/tasks specified in Section V.C.7.

Deliverables for Section V.C.7.

a) Quality control specifications that describe in detail all of the steps to be implemented to demonstrate to the Partnership that the final data are accurate

b) Quality control systems to verify the accuracy of the data processing, cleaning, and analyses.

Response

Pearson and ETS will be jointly responsible for the quality control systems to verify the accuracy of the psychometric services and the provision of high quality and accurate results. Pearson and ETS recognize that even the best-designed system, maintained by the most qualified and dedicated staff, can encounter problems and exceptions. A rigorous and reliable QC system should identify issues early, facilitate appropriate correction, and routinely provide measures of how things are running. Both Pearson and ETS have rigorous internal QC processes, and we will document our QC processes and plans for PARCC at several points in the project.

Quality Control Approach and Procedures

Pearson will provide a Quality Control Plan to PARCC shortly after the start of the contract. This plan will describe in detail the quality control procedures we will implement during the first operational administration. We will apply quality assurance steps and use mock data sets to verify all steps related to scanning and scoring paper/pencil tests, capturing and scoring computer-based tests, and the development for associated data files prior to the operational analyses.

Data extract files provided by Pearson will benefit from the quality controls surrounding our scoring process. We will configure the scoring system using the test maps and keys provided for the tests. We subject test maps to rigorous quality assurance checks to confirm the accuracy of data that contributes to item scoring. Our content group also performs independent key checks for operational test forms.

Once the system is configured our quality assurance group will verify that the selected responses entered by the student for an item correspond to the response recorded in the database, in pre-score and scored student data files. Scoring for selected-response items is verified against the keys and validations made for individual student's derived scores per level of the test. This process includes reviewing score-value-related fields such as raw scores, object scores, strand scores, performance levels, pass/fail indicators, attempt rules, and scale scores against the tables provided. This will result in data extract files that are useful for statistical analysis.

Our quality assurance group will perform acceptance testing on data extract files, as an independent means of quality control. These checks will include verifications that values are within defined ranges, checks to identify any missing values, and analyses to confirm the completeness of data records. These checks will be guided by a quality control plan, which will be provided to PARCC for review before processing.

Once data extracts are provided to ETS staff, they will conduct psychometric analyses is to verify the data are free from errors. ETS has a department of Data Quality Services (DQS) that runs comprehensive quality control checks on every data file we receive before psychometric analyses are begun. ETS psychometricians work with DQS to produce specifications designed to locate duplicate records and values out of range. A copy of the DQS specifications will be provided to PARCC.

Another step to providing quality psychometric analyses is to develop detailed statistical procedures. ETS will document our statistical procedures and QC processes for PARCC. For the PARCC operational analyses. We will work with Pearson to independently parallel process analyses using the same non-proprietary software. Psychometric analyses conducted at ETS undergo comprehensive quality checks by a team of psychometricians and data analysts. ETS will complete detailed checklists to establish that each of the statistical procedures is performed correctly for every analysis. Senior psychometric advisors and directors will review the results of calibrations and equating.

During the item analysis, DIF analysis, calibration, and equating process, ETS checks that the correct options within the analysis programs are chosen. ETS also checks the number of items, number of students with valid scores, IRT item parameter estimates, standard errors for the item parameter estimates, and compares preliminary operational statistics with those obtained during field testing. ETS psychometricians also perform detailed reviews of item test plots and statistics to investigate whether the data fit the model. ETS will check during the scaling process that the correct options for the analyses are used.

As a further psychometric quality control step, Pearson psychometricians will replicate IRT item calibrations, linkings, and scalings directly related to operational score reporting. This will include field test item calibrations, assuming use of these item parameters for pre-equating. Pearson psychometricians also will facilitate handoffs of data extracts to ETS and will support any investigations of anomalous data in the extracts provided by Pearson's IT group.

Independent Audit and Analysis of Psychometric Services

To provide further quality assurance, Pearson will contract with Measured Progress to provide an independent audit and analyses of the psychometric services performed on the project. The Measured Progress psychometric team brings together over 55 years of combined experience in the delivery of high-quality, large-scale assessments that meet or exceed the expectations set forth in the most recent versions of the Standards for Educational and Psychological Testing. Central to the Measured Progress philosophy is the implementation of data-driven solutions optimized to each client's unique requirements. The operational experience of Measured Progress encompasses implementation of a variety of item response theory approaches, ranging from one parameter to multi-parameter methods designed to accommodate innovative item types.

Each member of the Measured Progress psychometric team specializes in a unique area of the field of psychometrics, enabling the assignment of appropriate staff for a given assessment. And, although individuals may have separate areas of focus, this is a team that works together and shares research results and knowledge. This combined depth and breadth of expertise will provide a comprehensive, valid audit for PARCC that capitalizes on the most innovative thinking in the assessment community while grounding the assessments in valid, reliable measurement approaches.

For each of the major areas of psychometric services, Measured Progress will be the team that audits the psychometric work conducted by ETS.

Work will be conducted independently from ETS, and will occur at Measured Progress corporate headquarters in Dover, NH. Files will be transferred to Measured Progress psychometricians as the work nears completion.

Data files, artifacts, and supporting documentation will be placed on a secure FTP site hosted by Measured Progress. ETS will place copies of all files onto the Measured Progress site. ETS will have limited write-only access, and only key individuals will be given access to the Measured Progress site. All communications between ETS and Measured Progress will be documented and presented as part of the reporting of the audit process. .

Proposed Audit Process

Our strategy for auditing psychometric plans and documentation will follow a four step process:

1. Data handoff and communication of file specifications
2. Measured Progress review of inputs, outputs, and other artifacts
3. Conference calls involving PARCC, Pearson, and ETS to discuss critically needed changes/amendments – prior to the release of any reports
4. Report detailing results of audit process along with long term recommendations for improvement and refinement of work conducted by ETS

Upon completion of our audit process each year, we will submit a report with the following sections:

- Purpose of audit
- Handoffs (listing of all inputs, outputs, and artifacts used in audit process)
- Detailed description of the activities conducted during the audit
- Results of conference call with PARCC – including detailed descriptions of critically needed changes/amendments
- Detailed and comprehensive description of long-term recommendations from Measured Progress

For the time allocated for this task, Measured Progress will focus more attention in the initial year when the psychometric scales are developed and in subsequent years when the psychometric scales are equated. This is why more time is allocated for section V.C.1 compared to the other psychometric services that ETS will be providing. The tasks associated with section V.C.1 are critical to everything that is produced from the operational administration and the audit process will speak to accuracy and technical rigor.

Below are examples of some activities that would be conducted during the audit process to support activities in section V.C.1:

- Review of data cleaning rules to determine if properly executed
- CTT statistics for reasonableness (p-values, point bi-serial coefficients)
- Review of DIF statistics including review of items flagged with high levels of DIF
- Review of calibration details (see detailed specifications below)
- Review of the execution of equatings (including review of inputs, output, and artifacts created in the equating process)
- Review of scaling work (including proper execution of scaling rules such as rounding and truncation)
- Review of comparability studies between EOC and integrated high school mathematics tests
- Review of linking items – and methods used to remove equating items
- Review of sub-score to verify reliable and valid sub-score reporting
- Review of growth score calculations

High-level data quality Measured Progress psychometricians and data analysts will be involved in the audit process. Having both staff members involved in this part of the review process will verify that the data cleaning activities are meeting the expectations of PARCC and that the assessments and analyses are technically sound and designed to yield high quality data.

Measured Progress will use a similar process for all aspects of the audit process including our evaluation of data quality, data analysis, data forensics, technology & data requirements, field test analyses, retest analyses, and research studies.

Detailed Specifications

Below is an example of some of the specifications that Measured Progress would work with ETS on during the review process. In this example we have focused the presentation on the IRT calibration process. These specifications include:

- Inputs:
  - Student level data files (item x person level response data file)
  - Item list files – which would include the following type of specifications:
    - Item ID
    - Item purpose (e.g., scoring item, field test item, anchor item)
    - Item content information
    - Scoring details (maximum/minimum score points, weights)
  - Command files
    - IRT calibration program command files
    - Equating item parameters
    - Equating specifications
    - Equating software settings
- Outputs:
  - All IRT calibration program outputs
  - Equating output files
  - Equating item evaluations
  - Log of interventions used in analysis
- Artifacts
  - Item-model fit plots/statistics
  - Test characteristic curves and test information functions
  - Look-up tables and other scoring details

References

Agresti, A. (2002). *Categorical data analysis* (Vol. 359). Hoboken, NJ: John Wiley & Sons.

American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessments (APA) (1986). Guidelines for computer-based tests and interpretations. Washington, D.C.: American Psychological Association.

American Educational Research Associates (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (1999). Standards for educational and psychological testing. Washington, D.C.: AERA.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.

Angoff, W. H. (1974). The development of statistical indices for detecting cheaters. Journal of the American Statistical Association, 69, 44-49.

Atar, B. (2006). *Differential Item Functioning Analyses for Mixed Response Data using IRT Likelihood-Ration Test, Logistic Regression, and GLLAMM Procedures*. Unpublished Dissertation. Florida State University.

Baker, F. B., & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement*, *28*(2), 147-162.

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4): 42–51. doi: 10.1111/j.1745-3992.2009.00161.x

Betebenner, D. W., Van Iwaarden, A., Domingue, B., Shang, Y. (2013). *An R package for the calculation and visualization of student growth percentiles & percentile growth trajectories*. R package version 1.1-0.0.

Bock, R. D., & Zimowski, M. F. (1997). Multiple group IRT. In *Handbook of modern item response theory* (pp. 433-448). New York: Springer.

Braun, H., Ragosta, M., & Kaplan, B. (1986). *The predictive validity of the Scholastic Aptitude Test for disabled students*. ETS Research Report RR-86-38. Princeton, NJ: Educational Testing Service.

Bridgeman, B., Laitusis, C. C., & Cline, F. (2007) T*ime requirements for the different item types proposed for use in the revised SAT*. College Board Report No. 2007-35. New York: College Entrance Examination Board.

Bridgeman, B., Laitusis, C. C., & Cline, F. (2007) T*ime requirements for the different item types proposed for use in the revised SAT*. College Board Report No. 2007-35. New York: College Entrance Examination Board.

Bridgeman, B., & Lakin, J. (2012). *Predictive validity of the EXADEP ™ at the University of Puerto Rico*. ETS Research Memorandum RM-12-04. Princeton, NJ: Educational Testing Service.

Briggs, D. (2013). Making inferences about growth and value-added: Design issues for the PARCC consortium. . PARCC White Paper. Retrieved from http://www.parcconline.org/new-white-papers-available

Briggs, D. C., & Domingue, B. (2013). The gains from vertical scaling. Journal of Educational and Behavioral Statistics. Advance online publication. doi: 10.3102/1076998613508317

Buzick, H. M., & Jones, N. D. (under review). Using test Scores from students with disabilities in teacher evaluation.

Buzick, H. M., & Weeks, J. P. (in progress). *Validating the use of growth measures from statewide standards-based summative assessments for students with disabilities*. Institute of Education Sciences Grant (No. R324A120224).

Buzick, H., & Stone, E. (2011). *Recommendations for conducting differential item functioning (DIF) analyses for students with disabilities based on previous DIF studies*. ETS Research Report 11–34.

Buzick, H. M. & Stone, E. (2013, April). *A meta-analysis of research on the read aloud accommodation for K-12 students with disabilities*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Dorans, N.J. & Feigenbaum, M.D. (1994). Equating issues engendered by changes to the new SAT and PSAT/NMSQT. In I.M. Lawrence, N.J. Dorans, M.D. Feigenbaum, N.J. Feryok, A.P. Schmitt, & N.K. Wright (Eds.), Technical issues related to the introduction of the new SAT and PSAT/NMSQT (ETS Research Memorandum No. RM-94-10). Princeton, NJ: Educational Testing Service.

Cahalan-Laitusis, C., Mandinach, E., & Camara, W. (2002). *Predictive validity of SAT I: Reasoning Test for test takers with learning disabilities and extended time accommodations*. College Board Report 2002-5. New York: College Entrance Examination Board.

Castellano, K. E. (2011). *Unpacking student growth percentiles: Statistical properties of regression-based approaches with implications for student and school classifications*. Available from ProQuest Dissertations and Theses database. (UMI No. 3461371).

Castellano, K. E., & Ho, A. D. (2013a). *A practitioner's guide to growth models.* Washington, DC: Council of Chief State School Officers.

Castellano, K. E., & Ho, A. D. (2013b). Contrasting OLS and quantile regression approaches to student "growth" percentiles. *Journal of Educational and Behavioral Statistics*. doi: 10.3102/1076998611435413

Castellano, K. E., & Ho, A. D. (under review). Practical differences among aggregate-level conditional status metrics: From median student growth percentiles to value-added models.

Castellano, K. E., & Ho, A. D. (in progress). Predictive accuracy of projection models that afford growth-to-standard interpretations.

Chute, E. & Niederberger, M. (2012). Inquiry continues into cheating on PSSA tests. The Pittsburgh Post-Gazette. Retrieved from http://www.post-gazette.com/stories/news/education/inquiry-continues-into-cheating-on-pssa-tests-652513/

Cizek, G. J. (1999). Cheating on tests: How to do it, detect it, and prevent it. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Clark, J. M., Skorupski, W. P., Jirka, S., McBride, Y. Z., Wang, C., & Murphy, S. M. (2013). An investigation into statistical methods to identify unusual rates of item response agreement. Manuscript in preparation.

Clark, J. M., Skropuski, W. P., & Murphy, S. M. (2013). Using nonlinear regression to identify unusual performance level classification rates. Paper presented at the Conference on Statistical Detection of Potential Test Fraud, Madison, WI.

Davis, L.L., Strain-Seymour, E., & Gay, H. (2013). Testing on tablets: Part II of a series of usability studies on the use of tablets for K-12 assessment programs. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/Testing-on-Tablets-Part-II_formatted.pdf

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.

Dorans, N. J., & Feigenbaum, M. D. (1994). Equating issues engendered by changes to the SAT and PSAT/NMSQT. *Technical issues related to the introduction of the new SAT and PSAT/NMSQT* (ETS Research Memorandum 94-10). Princeton, NJ: Educational Testing Service.

Dorans, N. J. & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach*. (Research Report No. 91-47). Princeton, NJ: Educational Testing Service.

Donoghue, J. R., & Isham, S. P. (1998). A comparison of procedures to detect item parameter drift. *Applied Psychological Measurement*, *22*(1), 33-51.

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. British Journal of Mathematical and Statistical Psychology, 38, 67-86.

Fitzpatrick, A. R., Link, V. B., Yen, W. M., Burket, G. R., Ito, K., & Sykes, R. (1996). Scaling performance assessments: A comparison of one-parameter and two-parameter partial credit models. *Journal of Educational Measurement*, *33*(3), 291–314.

Glas, C. A. W. (2000). Item calibration and parameter drift. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 183-199). Norwell, MA: Kluwer Academic.

Guttman, L. (1944). A basis for scaling qualitative data. American Sociological Review, 9, 139-150.

Haberman, S. J. (2009). Use of generalized residuals to examine goodness of fit of item response models (ETS Research Report No. RR-09-15). *Princeton, NJ: ETS*.

Haberman, S.J. & Sinharay, S. (2010). “Reporting of subscores using multidimensional item response theory.” *Psychometrika*, 75(2): 209-227.

Haberman, S., Sinharay, S., & Chon, K. (2013). Assessing item fit for unidimensional item response theory models using residuals from estimated item response functions. *Psychometrika*, *78*(3), 417-440.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hanson, B. A., & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, *26*(1), 3-24.

Harnisch, D. L. & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. Journal of Educational Measurement, 18, 133-146.

Holland, P. W. (2002). Two measures of change in the gaps between CDFs of test score distributions. *Journal of Educational and Behavioral Statistics*, *27*(1), 3-17.

Holland, P. W. (1996). Assessing unusual agreement between the incorrect answers of two examinees using the K-index: Statistical theory and empirical support (ETS Tech. Rep. No. 96–4). Princeton, NJ: Educational Testing Service.

Holland, P., & Rubin, D. (1982). *Test equating*. New York: Academic Press.

Holland, P. W., & Thayer, S. T. (1988). Differential item performance and the Mantel-Haeszel procedure. In H. Wainer & H. Braun (Eds.) Test Validity (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum.

Ho, A. D. (2007). Discrepancies between score trends from NAEP and state tests:

Howell, D. C. (2010). *Fundamental statistics for the behavioral sciences*. CengageBrain. com.

Ito, K., Sykes, R. C., & Yao, L. (2008). Concurrent and separate grade-groups linking procedures for vertical scaling. *Applied Measurement in Education*, *21*(3), 187-206.

Jacob, B. A. & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. The Quarterly Journal of Economics, 118, 843-877.

Jones, D., & Ragosta, M. (1982). *Predictive validity of the SAT for two handicapped groups: The deaf and the learning disabled*. ETS Research Report RR-82-9. Princeton, NJ: Educational Testing Service.

Kahraman, N., & Thompson, T. (2011). Relating Unidimensional IRT Parameters to a Multidimensional Response Space: A Review of Two Alternative Projection IRT Models for Scoring Subscales. *Journal of Educational Measurement*, *48*(2), 146-164.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. Applied Measurement in Education, 16, 277-298.

Kim, S. H., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied psychological measurement*, *22*(2), 131-143.

Keng, L., Kong, X. J., Bleil, B. (2011, April). Does Size Matter? A Study on the Use of Netbooks in K-12 Assessments. Paper presented at Annual Meeting of the American Educational Research Association, New Orleans, LA. Retrieved from: http://www.pearsonassessments.com/hai/images/PDF/AERA-Netbooks_%20K-12_Assessments.pdf

Kolen, M. J., & Brennan, R. L. (2004). Test equating, linking, and scaling: Methods and practices. New York: Springer-Verlag.

Laitusis, C. C. (2010). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education, 23*(2), 153–167.

Laitusis, C., Buzick, H., Stone, E., Hansen, E. & Hakkinen, M. (2012). Literature Review of Testing Accommodations and Accessibility Tools. Commissioned Report for the Smarter Balanced Assessment Consortium. See http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/08/Smarter-Balanced-Students-with-Disabilities-Literature-Review.pdf

Lakin, J. M., & Young, J. W. (2013). Evaluating growth for ELL students: Implications for accountability policies. *Educational Measurement: Issues and Practice, 32*(3), 11-26.

Luecht, R.M. & Camara, W.J. (2011). Evidence and design implications required to support comparability claims. Retrieved from http://www.parcconline.org/sites/parcc/files/PARCCWhitePaperRLuechtWCamara.pdf

Mandinach, E. B., Bridgeman, B., Cahalan (Laitusis), C., & Trapani, C. (2005). *The impact of extended time on SAT I Reasoning test performance*. College Board Report 2005–8. New York: College Entrance Examination Board.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58,* 690-700.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719-748.

Mislevy, R. J. (1987). Recent developments in item response theory with implications for teacher certification. *Review of Research in Education*, *14*, 239-275.

National Council on Measurement in Education (2012). Testing and data integrity in the administration of statewide student assessment programs. Retrieved from http://ncme.org/default/assets/File/Committee%20Docs/Test%20Score%20Integrity/Test%20Integrity-NCME%20Endorsed%20(2012%20FINAL).pdf

New Jersey Department of Education (2012). 2011 Erasure Analysis Report: New Jersey Assessment of Skills and Knowledge (NJ ASK 3-8). Retrieved from http://education.state.nj.us/broadcasts/2012/AUG/16/7683/2011%20NJ%20ASK%20Erasure%20Analysis%20Report%20FINAL.PDF

Pell, M. B. (2012). More cheating scandals inevitable, as states can't ensure test integrity. The Atlanta Journal-Constitution. Retrieved from http://www.ajc.com/news/news/more-cheating-scandals-inevitable-as-states-cant-e/nSPqj/

Puhan, G., Moses, T. P., Grant, M. C., & McHale, F. (2009). Small-Sample Equating Using a Single-Group Nearly Equivalent Test (SiGNET) Design. *Journal of Educational Measurement*, *46*(3), 344-362.

R Core Team (2013). *R: A language and environment for statistical computing*. [computer software.] Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org.

Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement*, 25(3), 193-203.

Reckase, M. D., Martineau, J. A., & Kim, J.-P. (2000). *A vector approach to determining the number of dimensions needed to represent a set of variables.* Paper Presented at the Annual Meeting of the Psychometric Society. Vancouver, Canada.

Sijtsma, K. (1986). A coefficient of deviance of response patterns. Kwantitatieve Methoden, 7, 131-145.

Smith, R.L., Wang, M.M., Wingersky, M., & Zhao, C. (2001). Monitoring items for changes in performance in computerized adaptive tests. Paper presented at the annual conference of the National Council on Measurement in Education, Seattle, Washington.

Sotaridona, L. S. & Meijer, R. R. (2002). Statistical properties of the K-index for detecting answer copying. Journal of Educational Measurement, 39, 115-132.

Sotaridona, L. S. & Meijer, R. R. (2003). Two new statistics to detect answer copying. Journal of Educational Measurement, 40, 53-69.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.

Stone, E., Cook, L., Laitusis, C. C., & Cline, F. (2010). Using Differential Item Functioning to investigate the impact of testing accommodations on an English language arts assessment for students who are blind or visually impaired. *Applied Measurement in Education, 23*(2), 132–152.

Stone, E., Cook, L.L., & Laitusis, C.C. (2013). *Evaluation of a condition-adaptive test of reading comprehension for students with reading-based learning disabilities*. ETS Research Report RR-13-20. Princeton, NJ: Educational Testing Service.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, 589-617.

Strain-Seymour, E., Craft, J., Davis, L.L, & Elbom, J. (2013). Testing on tablets: Part I of a series of usability studies on the use of tablets for K-12 assessment programs. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/Testing-on-Tablets-PartI.pdf.

Swaminathan, H., & Rogers, H.J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27,* 361-370.

Wainer, H., & Thissen, D. (2001). True score theory: The traditional method. *Test scoring*, 23-72. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Way, W. D., Davis, L. L., & Fitzpatrick, S. (2006, April). Score comparability of online and paper administrations of the Texas Assessment of Knowledge and Skills. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

Way, W. D., Um, K., Lin, C., & McClarty, K. L. (2007, April). An evaluation of a matched samples method for assessing the comparability of online and paper test performance. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Wollack, J. A. (1997). A nominal response model approach for detecting answer copying. Applied Psychological Measurement, 21, 307-320.

Wollack, J. A. (2006). Simultaneous use of multiple answer copying indexes to improve detection rates. Applied Measurement in Education, 19, 265-288.

Yen, W. M., & Ferrara, S. (1997). The Maryland School Performance Assessment Program: Performance assessments with psychometric quality suitable for high-stakes usage. *Educational and Psychological Measurement*, *57*(1), 60–84.

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential ItemFunctioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores.* Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

PARCC

Component 2: Assessment Administration

V.B.1 Technology Requirements

Requirement

Response Requirements for Section V.B.1.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.1

b) For contingency purposes, PARCC would like the Offeror to provide the following cost options:

   i. Hosting, maintenance, and updates for PARCC's Data Management and Reporting Components.

   ii. Contractor-provided Assessment Content, Assessment Delivery, and Shared Service for years one through four. For this option, the Contractor is not required to follow PARCC's interoperability requirements for data exchanges between Contractor supplied components. The Contractor would be expected to follow PARCC's interoperability requirements for data exchanges (item/student/organization) to/from the Contractor's and PARCC's data warehouse and reporting components. The Contractor shall identify areas where meeting PARCC's requirements, would delay or prevent a successful implementation in year one.

Deliverables for Section V.B.1.

a) Forms Management Metadata

b) Administrative/Statistical Metadata

c) Student and Organizational Registration Data d) Student Response Data (In Development)

e) Scoring I Results Data (In Development)

Response

For its operational assessments, PARCC requires a partner who understands the wider landscape into which the result of this RFP must fit. The technology philosophy and systems architecture need to be designed to support states in administering next-generation assessments in a way that empowers the Partnership's long-term sustainability.

We recognize the need to work in a multiple-vendor setting and to make the interaction and process is as smooth as possible. There are many "moving parts" that must be understood and coordinated. To this end, we anticipate clear communication and cooperation from our staff as well as that from other vendors, such as the Data Warehousing and Report Design

Technology Bundle provider, who will commit to providing technical training and guidance on their delivered components.

This training will be focused on, but not limited to, the following areas:

- Component administration
- Component configuration and architecture
- Component data architecture and migration procedures
- Component integration management
- Component testing
- Component deployment processes
- Component upgrades

Data Privacy and Security

Pearson works with providers to create secure hosting environments that protect network assets and information from unauthorized access or operations disruptions. We maintain a Security Policy and Requirements document outlining strict procedures for the physical security of hosting facilities.

Protection from network-based threats is as important as physical security. Pearson engineers, in conjunction with data center staff, deploy, manage, and monitor the security of Pearson systems. We employ a variety of security technologies and tools in the computing and network environments. External penetration testing and security scanning are routinely performed to verify that our systems are adequately hardened and protected form security threats.

As requested, additional information about security is provided in other sections of our response. We will comply with federal laws and PARCC policies for data privacy and security that include how data are accessed, stored, and exchanged, and employees working on this project will be trained on PARCC security protocols.

Accessibility and Fairness

PARCC requires an online testing system that provides each student the opportunity to test. User accessibility guidelines play a key role in defining solutions for accessibility, which include WCAG and Section 508 standards. Pearson strives to adhere to WCAG and Section 508 standards.

Occasionally, requirements such as comparability across devices and how different devices perform or support accessibility may prevent complete compliance. The World Wide Web Consortium (W3C) also recognizes the need for flexibility, and created three conformance levels in WCAG 2.0, stating, “It is not recommended that Level AAA conformance be required as a general policy because it is not possible to satisfy all Level AAA Success Criteria for some content" (source: http://www.w3.org/TR/WCAG20/#conformance-reqs).

To the extent possible, Pearson works to bring its web-based products into compliance with these standards.

The Section 508 Technical Standards that apply primarily cover software usability specifications for people with vision impairments and software compatibility with adaptive equipment. We acknowledge the Section 508 standards and strive to meet them. If a conflict emerges between meeting the standard and the overall goals of assessment, we will work with PARCC to resolve the conflict.

For our delivery system, we use the IMS Global Accessible Portable Item Protocol (APIP) open interoperability standard for assessment content and meta-data encoding. APIP provides a robust content tagging and meta-data vocabulary specifically designed for accessible assessment content encoding. APIP content tagging and extensions are designed to support a wide range of disabilities and capabilities within the delivery system.

Pearson continues to progress in our accessibility and accommodation capabilities. Pearson's proprietary test delivery platform will provide the following benefits and features:

- **Interoperability.** Accessibility data is encoded in the item in a way that can be moved across delivery systems (APIP).
- **Flexibility.** Features can be enabled individually, as desired, without including features that should not be allowed, such as specifying which words can be defined or translated or determining how equations are read aloud.
- **Lower Cost and Simpler Logistics.** Some assistive technology purchases may become unnecessary as a student can test on any approved device
- **Security.** Software running in the background, which can be a security violation, is not needed.
- **Simple Reviews**. Customers can review items with the accommodations in place and request corrections or adjustments if needed.
- **Standardization.** Accommodations implementation follows APIP.
- **Customization.** Accommodations can be customized using a Personal Needs and Preferences (PNP) profile, as described by APIP.

Pearson has engaged with the following businesses working on WCAG compliance and accessibility, and is in the process of integrating recommendations for accessibility in our existing proprietary platforms:

- American Printing House for the Blind, Louisville, KY (http://www.aph.org)
- gh, Lafayette, IN (http://www.gh-accessibility.com)

- Knowbility, Austin, TX (http://www.knowbility.org)
- TextHelp, Woburn, MA (http://www.texthelp.com)
- CAST, Wakefield, MA (http://www.cast.org)

PARCC's goals depend on a system that complies with accessibility standards whenever possible while maintaining the security and comparability of assessments. Pearson looks beyond general accessibility standards to provide accommodations specifically designed for students testing online.

APIP currently provides the most complete set of standards for accommodations in high-stakes assessments. We play a leadership role in defining and supporting open standards like APIP, and we continue to research how accommodations affect student performance.

Compatibility with School Technology Infrastructure

PARCC requires a solution that is designed to provide optimal performance in high-technology capability settings that have current generation computers and large bandwidth networks, but still function without sacrificing performance in low-technology capability settings.

Pearson recognizes the PARCC core principle that includes a "device agnostic" approach to assessment content and assessment technology development and will deliver components that are designed to function comparably across a range of devices using commonly deployed web browsers, including desktops, laptops, netbooks, and tablets (9.5" or larger) running Windows, Mac, Linux, Apple iOS, Android, and Chrome operating systems as detailed in the PARCC Technology Guidelines and Technical Specifications.

Interoperability

Pearson supports PARCC's commitment to open technology interoperability standards. Pearson's next generation systems are being built using open-source technologies and open interoperability standards as a core principle. When assessment solution providers share common standards, their assessments, assessment items, and assessment data become highly portable across systems. Pearson's next generation assessment systems meet the guidelines outlined by the US Department of Education in the Race to the Top Assessment program.

The specific interoperability standards that are core to the Pearson assessment platform are the QTI and APIP standards. PARCC content is being encoded to QTI and APIP specifications. This includes the base item content as well as the item scoring and assessment test information. Any system capable of consuming this format should be able to efficiently receive and process PARCC assessments. Because these standards allow for flexibility in implementation, any system receiving PARCC content will need to adhere to the PARCC content profile for QTI and APIP.

Not only is Pearson's solution based on open interoperability standards, it is designed to scale to support a wide range of technologies and platforms. Our experience has helped us design platforms that operate efficiently in the most modern and large bandwidth networks as well as in low technology and more limited bandwidth settings. Our test delivery solution is using presentation strategies that place "mobile first." From that strong foundation we can easily scale to larger devices and platforms. We support PARCC's vision for an assessment platform that is flexible, scalable, and supports the technology environments found in schools and districts today.

Interoperability Conformance and Validation Data Privacy and Security

Incompatible systems can increase development costs, cause delays, and reduce functionality. Pearson works closely with standards organizations to develop and enhance their data and content standards. We have based our next generation systems on open interoperability standards that enable system interfaces to exchange data and content in a standard way.

Standards have limitations and often provide for ways to extend the standard with proprietary implementations. With the industry moving toward more technology-enhanced assessments, our guiding principles are to work within the standard framework and document extensions that may be used to implement innovative functionality not natively supported by the standard.

To continue improving interoperability standards for content and data, Pearson maintains a leadership role in defining and supporting XML, APIP, QTI, and SIF to provide new opportunities for PARCC to reduce costs while increasing the instructional benefits of assessments. Isolated, non-compliant testing platforms cannot keep pace with changing regulatory demands or provide the efficiency of interoperable systems.

Pearson's system shall conform to applicable industry-recognized, open-licensed interoperability standards including Assessment Interoperability Framework 1.0, Common Education Data Standards 3.0, QTI 2.1 and APIP 1.0 standards, and any extension of such standards which are required to support PARCC's assessment items, assessment and results data, accessibility, student data, and APIs. Where PARCC has identified specific standards, schemata, and controlled vocabularies, Pearson systems shall conform to PARCC interoperability guidelines. Where PARCC is in the process of developing guidelines around the application of existing standards, Pearson shall participate. Where open-licensed standards do not exist or are inadequate to support PARCC's assessment items, assessment and results data, accessibility, or student data, Pearson shall work collaboratively with PARCC's Interoperability Services contractor (procured separately) to propose existing Pearson solutions or develop appropriate new extensions for adoption by the applicable standards development organizations. Pearson shall work collaboratively with PARCC's

Interoperability Services contractor to represent PARCC's needs in the community participation processes of relevant standards setting bodies.

We can draw upon extensive in-house experience. For example, Wayne Ostler, Vice President of Digital Content and Measurement Systems; Jason Craft, Principal Software Engineer; and Michelle Richard, Manager Content Encoding and Transformation services, actively participate on the APIP Working Group (known as the APMG) and are part of the management team that will provide executive leadership and management to the project.

PARCC Vendor Responsibilities

For the delivery of the PARCC Operational Assessments program, Pearson will work with other vendors, across PARCC contracts as described below:

- **The Vendor for PARCC's Data Warehouse and Reporting Components.** As part of the Operational Assessments scope of work, Pearson will be responsible for management and administration of the work that uses these components.
- **The Vendor for the PARCC Resource Center.** As part of the Operational Assessments scope of work, Pearson will be responsible for making released items and test forms available through the Resource Center RFP.
- **The vendor for the PARCC item bank.** As part of the operational assessments scope of work, Pearson will be responsible for delivering all items and test forms to the PARCC item bank in the PARCC QTI-APIP item format.
- **The Vendor for PARCC Technology Operations.** As part of the Operational Assessments scope of work, Pearson will be responsible for managing, coordinating, and supporting the customer-facing administration activities using the Partnership-owned technology components.

The delivery platform for the content being developed for the Diagnostic Assessments, K-1 Formative Assessments, and Speaking and Listening Assessments will be procured separately and is outside of the scope of work for this contract.

Additionally, Pearson's role for the Operational Assessments contract will include providing the remaining components. Finally, as Pearson-owned technologies employed for PARCC operational administration are updated, Pearson will make the updated versions of the systems it uses to provide service available for PARCC access pending PARCC approval, at no additional cost to PARCC, conditioned on the requirement that such updated versions of the systems would not require material modifications to the PARCC assessment items and forms for proper rendering within such updated versions of the systems.

Assessment Delivery Platforms

Pearson will provide access to its proprietary administrative portal (PearsonAccess) and computer-based testing platform (TestNav) for all four years of the contract. Should PARCC decide to use a different platform in the last year of the contract, PARCC will notify Pearson by September 1, 2016 for the assessments to be administered in the 2017-2018 school year. If that should happen, revised pricing will be negotiated. Conditioned upon the third party testing platform's ability to process enrollment data in the appropriate interoperable format, PearsonAccess will collect enrollment information for both paper and online testers. When updates and revisions to these platforms are made, Pearson will provide PARCC with the option of employing the updated version of these platforms. In addition, as PARCC identifies bugs in these platforms that impact Pearson's ability to comply with the functionality requirements under the contract, Pearson will remedy such bugs as needed to comply with the terms of the contract. In the event PARCC identifies preferential improvements or enhancements which may create efficiencies, PARCC may request such enhancements through the Product Review Board, and Pearson will consider incorporating these improvements as contemplated by the Product Review Board.

Shared Services

Pearson will work with other Partnership vendors to use the following shared technology functionalities to complete the activities and deliverables required to fulfill the requirements of this RFP:

- Authentication/user identity management
- Logging and audit
- System monitoring and alerting
- Common ID system

Item Bank

During the first operational year in 2014–15, Pearson will complete summative test form construction activities by working with assessment items that will be stored in the current Pearson item bank as part of our Assessment Administration Contract Amendment 3 scope of work. All ELA and Math content will not be in the ABBI repository to support form construction in Fall of 2014, so the current Pearson item bank will be used. In Fall of 2015, we will provide all content assets directly from the ABBI repository to support the form construction and publishing activities. In Fall of 2016, Pearson will have all test map form building, management and publishing functions available in ABBI to support the form construction and publishing activities directly in the ABBI repository.

Interoperability Requirements

Open standard formats are important for interoperability between discrete components of PARCC's Assessment System. Pearson is committed to supporting and using open standard formats where they are most useful and necessary, for functions such as data transfers to and from discrete components.

For PARCC Operational Assessments, Pearson will employ the PARCC interoperability standards and guidelines for the following:

- Item content
- Forms management metadata
- Administrative/statistical metadata
- Student and organizational registration data
- Student Personal Needs Preference data
- Student response data
- Scoring/results data

Pearson will work with PARCC to agree on standards and how they will be used before proceeding with data activities and deliverables.

Integration Requirements

The PARCC Data Warehousing and Reporting vendor will collaborate with our team to transition management activities to Pearson. We will be responsible for managing the integration of assessment components including Pearson-supplied components and PARCC-supplied components for the duration of the contract.

Pearson will establish a team to work collaboratively with the Partnership manager and other Partnership vendors to oversee quality control during the transfer of data from the test delivery platform data store (including registration and score data) into the Data Management and Reporting System being built the data warehouse and reporting development vendor

Tech Systems Cost Option

The Base Cost Proposal includes costs associated with student enrollment, item banking, and computer-based testing (including hosting and maintenance) for all four years of the contract. Separately, PARCC is procuring the development of the data warehouse and reporting component and the Partnership Resource Center.

Requirement

V.B.2.A. Scope

Response Requirements for Section V.B.2.A.

Response

Beginning in the 2014–15 school year, the PARCC Summative Assessments in English Language Arts/literacy and mathematics will be available for states to administer to students in grades 3 through 8 and high school.

Students will take two assessment components: the Performance Based Assessment (PBA) and the End-of-Year (EOY) Assessment in both content areas in grades 3-8. At high school students will take both the PBA and EOY components for the ELA and/or mathematics test(s) they are registered for, which will be based on each state's high school assessment adoption plan. The assessments will be delivered during one of the administration windows outlined below each school year. (As noted in the RFP, block scheduling only applies to high school ELA/literacy courses and mathematics, algebra I, geometry, algebra II, and integrated mathematics I through III.) Districts will have a four-week period for each summative assessment component (PBA and EOY) in which to complete testing. For the Fall/Winter 2014 Block the testing window may be less than four weeks as mutually agreed on by PARCC and Pearson. Six of the nine high school assessments will be available for the Fall/Winter 2014 Block administration in paper-based format only. The integrated math tests and online testing will be available starting with the Traditional Spring administration.

Administration Window	Approximate Timing
Fall/Winter Block	PBA: November-December EOY: December-January
Traditional Year	PBA: Mid-February-Mid April EOY: Early April-Late May
Spring Block	PBA: Early April-Mid May EOY: Early May-Mid June

See the attached Materials List (revised 04/06/14) for additional assumptions about materials that will be provided for each administration and specifications for the materials including document sizes for answer document and test booklet.. Annual student volumes beginning in year 2 will also include retest administrations (administered during the regular testing windows). Individual states can work with Pearson as needed to negotiate summer retesting as needed, which is not included in the base scope of work but will be included in the annual student volumes count.

Requirement

V.B.2.B. Security

Response Requirements for Section V.B.2.B.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.B

Deliverables for Section V.B.2.B.

a) Deliver Security Plan

Response

The security plans proposed for the PARCC Operational Assessments are part of Pearson's standard security and technology policy as described below. By aligning our security policies and principles with the ISO/IEC 27000 set of standards, Pearson's protection strategies adhere to internationally recognized standards and best practices in security.

During our many years of processing confidential information, we have developed rigorous standards to secure confidential data throughout its lifecycle. This helps us meet regulatory requirements for secure handling of confidential data set forth by federal statutes such as the Federal Educational Rights and Privacy Act (FERPA). The following provides an overview of the administrative, logical, and physical security controls that Pearson employs to protect our customers and their data.

Overview of Administrative Security Controls

Global Information Security Policy

Pearson's security efforts begin with executive direction and support, which is formalized in the Global Information Security Policy. The policy is an ISO27000-based document providing the baseline expectations of all Pearson operational companies. Key security considerations such as protection of payment card data, personally identifiable information (PII), and other sensitive customer data are addressed, requiring strong control processes and technologies to meet not only Pearson's requirements, but external regulation and contractual obligations as well.

The policy is managed through the Corporate Information Security Department, Office of the Chief Security Officer (CSO). The group provides policy and compliance efforts through appointed data security or privacy officers located in strategic business locations. These individuals assist each business unit by providing overall security services, including consistent policy guidance, evaluation of technical controls for compliance, resources for security awareness training, support for internal and external audit activity, and remediation efforts. The group is also staffed with dedicated security analysts who manage Pearson's security applications, security systems, forensic analysis, and incident response and containment, including root cause analysis and post-incident remediation efforts.

Security Awareness Program

Pearson's Security Awareness Program is an ongoing effort providing guidance to every employee so they understand Pearson's Security Policies, their individual responsibilities for compliance, and how their behaviors affect Pearson's ability to protect systems and data. The core of these efforts is built on the well-known principles of Confidentiality, Integrity, and Availability (CIA). Security awareness begins immediately with orientation for new hires.

Training covers acceptable use of Pearson systems and fundamental best practices, such as creating strong passwords, proper use of email and Internet access, and responsibility to report security risks. Web-based training modules are available, and all employees are required to complete the course within 30 days of hire and thereafter are routinely asked to refresh their knowledge. Via these modules, employees discover how each individual plays a significant role in protecting Pearson's information assets.

For example, employees are offered simple methods for choosing a secure password and preventing viruses, hoaxes, and intruders from accessing sensitive data and computing systems.

Specific online courses covering Payment Card Industry (PCI) compliance and PII have also been developed. An annual refresher course including components of all of these control areas is also provided in compliance with Pearson's awareness training requirements.

Continued awareness is maintained through security newsletters, business unit specific communications with targeted content for a relevant concern, protecting PII, payment card data, and notices of new policy revisions or additions.

Incident Management

Identifying and responding to security incidents is an important part of Pearson's security program. The Pearson Incident Management and Communication Plan defines detailed roles and responsibilities initiated when actual or potential security incidents are identified. This plan follows industry best practices to provide quick response, effective isolation and containment, thorough root cause analysis, and appropriate remediation.

Forensic analysis is also performed when necessary to provide detailed evidence for root cause analysis and any possible legal action. Pearson personnel are required to report any security incidents, violations of policy, or potential risks when they encounter them. They are provided with guidelines, points of contact, and overall expectations through Pearson policy, new-hire orientation training, and online security awareness courses.

Overview of Logical Security Controls

Vulnerability Management and Network Security

Pearson Security and Pearson Technology maintain an ongoing vulnerability management and network security program to confirm that technological solutions and consistent processes are in place to address the challenge of providing secure services.

Antivirus Software

Pearson uses centrally managed and updated client side antivirus software on all computers determined to be susceptible to virus infection. Updated signature files are distributed from a central management system. Users logging into the network are automatically kept up-to-date. In the event of an emergency, updated signature files can be manually "pushed" to address immediate risks.

Web Filtering at Internet Gateways

Pearson uses Internet web filtering software to prevent access to certain Internet content. Sites that are explicitly non-business related or known to be sources of malware and virus infection are blocked.

Extensive Email Filtering

Pearson has extensive email controls to prevent infected email, malicious attachments, and phishing attempts from reaching our internal network. The Pearson network protects against unwanted, unauthorized email, further protecting the systems and data it supports.

Intrusion Prevention\Detection

A state-of-the-art IDS\IPS system protects Pearson's publicly facing gateways and specific internal network transit points. Malicious content is dropped and potentially malicious behavior is identified with logging and analysis functionality. Custom signature files can be deployed to protect against zero-day exploits and manufacturer updates are installed upon receipt. This system provides monitoring 24 hours a day, seven days a week, and 365 days a year, along with real-time alerting to the security staff for immediate response. The IDS\IPS system also aids in identifying network activity that could indicate inappropriate behavior or infected computers.

Firewalls

The internal Pearson network is isolated from the public Internet by a layered firewall approach that creates a secure LAN environment with web, application, and database DMZs for publicly accessible services. Packet inspection firewall devices are used with only business-required ports and protocols enabled across the network boundaries. All firewall traffic also traverses an in-line IDS\IPS system to protect against network-based attacks, such as Denial of Service (DOS) and other known malicious traffic (see Vulnerability Management). Internal LAN segments are structured to improve performance and controlled access.

Control of Administrative Rights

Pearson requires tightly controlled and documented administrative rights. This strategy is designed to control the secure configurations deployed on workstations and laptops. Only staff members with specific job requirements and approval from Pearson Security are granted local administrative rights.

Full Disk Encryption for Laptops

Pearson has deployed a full disk encryption solution in its mobile computing environment to protect all data in the event that a laptop computer is lost or stolen.

Patch Management Processes

Pearson Technology employs a patch management process in both the production server environment and the desktop user environment. Critical patches are deployed within 30 days of vendor availability, and all other appropriate patches are installed quarterly. Scans are run to determine operating system status and to report the success or failure of scheduled system upgrades.

These tools, processes, and strategies provide a robust “defense in depth” vulnerability management and network security strategy for all applications and services running on Pearson systems.

Server and System Security

Standard Build

Pearson Technology Services (PTS) supports a wide variety of computer processing platforms, including the following:

- Sun Solaris on Sunfire and Dell/Intel servers
- Red Hat Linux Enterprise Server on Dell servers
- AIX on IBM P-Series Servers
- HP-UX on HP Servers
- Windows on Dell servers
- IBM zOS on IBM Z Series Mainframe

Pearson keeps standard builds current by monitoring emerging best practices for supported operating systems—Linux, Windows, and Sun, for example—and analyzing how developments can benefit our environments.

Configuration and Patch Management

One of the tools Pearson uses to monitor and manage Data Center operations is called System Insight Manager (SIM). This tool, built internally by Pearson, centralizes significant amounts of disparate data into a graphical, summary-based dashboard view.

Using SIM, Pearson Data Center personnel have comprehensive data on backup and recovery, maintenance schedules, patch management, obsolescence planning, system policy compliance, security and authorization, contact information and issue resolution, and performance monitoring. Our patching process and insight has regularly gained accolades from third-party auditors.

Logging, Auditing, and Monitoring

Pearson tracks access to/changes made to network components in a variety of ways, including:

- Access-logging features on all applicable components
- Logs written to a centralized management systems where they are backed up and protected from unauthorized access or modification

- Baseline configuration files maintained on the centralized management server; comparison scripts are executed weekly against currently running configurations to verify that all changes are recognized/authorized

Pearson provides a security management audit trail by enabling operating system (Windows, AIX, Solaris, HPUX, Redhat) audit features. The operating system records each attempted user-resource interaction with the password and user ID, which permits the audit of each individual's actions. Audit trails provide records for each activity in the system, including actions such as when a user attempts to read, modify, add, create, or delete information as well as attempting actions that require administrator-privileges.

For each recorded event, the audit record identifies the following:

- Date and time of event
- User
- Type of event
- Success or failure of the event
- Name of the object being used or deleted
- Log-on and log-off activity
- Database administrator activities
- Database modification activities and reasons for modification

An automated host-based assessment tool mines system logs on a nightly basis. Pearson Help Desk staff review these logs on a weekly basis. Any anomalies or suspicious behavior are immediately reported to support staff for review.

Since system logs are reviewed on a regular basis and can become unmanageable over time, logs are rolled to tape backup and stored offsite on a 90-day basis. This provides an audit trail of system and individual access attempts that can assist in forensic and incident handling processes.

The Pearson Help Desk reviews the logs weekly with primary emphasis on unsuccessful access events (computer security incidents). Access to operation system security features is granted only to administrative-level operators, protecting these audit logs from unauthorized reads or modification. Backup procedures help mitigate the risk of losing an audit trail from remote tampering or system failure.

Data Security/Encryption

Encryption

Pearson uses encryption protocols in many of its standard services, and can provide specific solutions based upon program requirements. Some of the more relevant uses of encryption are the following:

- **Secure FTP**—encrypts sensitive data between Pearson and our customers
- **HTTPS**—encrypts with Secure Socket Layer (SSL) on websites to protect all user transactions
- **Encrypted Backup Tapes**—encrypts backup tapes generated for offsite storage using hardware layer encryption to prevent data exposure through loss or theft of offsite media
- **Virtual Private Networks (VPN)**—provides secure remote access to the Pearson network and secure point-to-point network connections when necessary
- **Database Encryption**—protects data at rest from unauthorized access by implementing the correct encryption protocol for any customer data

Working with our customers, Pearson can provide the technical resources necessary to implement the correct encryption protocol for any program.

Data Destruction and Chain of Custody

Two critical components of end-to-end data security are proper disposal of data at end-of-life and secure transfer of media containing sensitive data in circumstances where the data cannot be removed. Pearson provides policy guidance through the Electronic Media Disposal Policy, which is implemented with specific processes for degaussing or securely wiping all media before it is removed from secure Pearson Technology locations, such as the Iowa City Data Center.

In the instances when systems or media must be transported with resident data, a strict chain-of-custody process is followed to confirm proper handling and protection of all data. This process requires written signatures at each stage of the transfer process, as well as acknowledgement of the Data Security Officer.

Secure Access Controls

Although encryption of data is a critical control, maintaining secure access to data is just as important.

For internal facing systems and applications, Pearson authorizes access to networks, systems, and applications based on business need to know, or the least privilege necessary to perform stated job duties. Any elevated rights not granted by virtue of job responsibilities at time of hire must be approved by management, and in certain instances, the Information Security Group (such as administrative rights).

For externally facing systems and applications, access to specific data is determined based on user roles and permission requirements established by our customers. Our staff will work with the customer to define role templates, with varying levels of security, for the customer's different user types. This process is typically completed prior to implementation and our goal is to enable our customer's to manage access to their data in a self-service model.

Network Access Controls

Pearson policy requires system access to be granted with a unique user ID. Every employee is provided with a unique domain account and a secure one-time password. To safeguard confidentiality, passwords must be changed upon first logon. For strong passwords, Pearson stipulates a minimum password length as well as complexity requirements (one number, upper case, special characters, and so on).

Accounts are locked out after five failed login attempts and can only be reset by calling the Pearson Help Desk. Passwords must be changed every 60 days, and cannot be reused within one year. A screen lock is also enabled by Domain Security Policy to activate after 15 minutes of inactivity. All accounts are locked at time of termination. To verify compliance, domain controls identify and lock all accounts that have been idle for more than 30 days. If they are still unused after 90 days, the account is deleted.

Application Access Controls

At a minimum, applications hosted on Pearson systems are required to provide the same level of security as network access would require. Many applications use the Pearson Active Directory to authenticate users with their domain credentials. This complies with corporate policy and provides easy use for the end user by keeping additional logon IDs to a minimum. If for any reason a centralized user database such as Active Directory cannot be used, unique application IDs and passwords following the same corporate requirements are assigned. Pearson's IBM z/OS Mainframe computer uses RACF (Resource Access Control Facility) for strict access control to the resources hosted on the system.

Remote Access Controls

Employees authorized to work remotely must access the Pearson network through a secure network connection. The authentication is tied to their domain account ID. Specific VPN profiles are constructed to isolate remote access to only the resources required by the individual user. When working remotely, the host-based firewall is enabled, protecting the remote client and the internal Pearson network from external compromise. All of these controls, along with full disk encryption on all Pearson laptop computers, provide a secure environment for Pearson's remote workforce.

System Access Controls

Direct access to Pearson systems is strictly controlled, both physically (see Data Center Security) and logically. Access is limited to a small group of authorized administrators, programmers, and database administrators (DBAs). Activity performed with root privileges can only be executed after authenticating with an individual unique ID.

Separation of duties between system administrators, application developers, application DBAs, and system DBAs provides clear definition of responsibility and prevents an individual from attaining the access rights that would allow for easy compromise. Administrative-level accounts are reviewed quarterly and passwords on shared accounts, such as root in the Unix environment, are promptly changed when administrative members leave the teams.

Overview of Physical Security Controls Facility Access Controls

All Pearson facilities are secure and closed to the general public. Physical access to each facility is controlled by an access card reading system. Employees are required to wear a company-issued photo ID badge. This Security Identification Badge is to be worn in unobstructed view at all times on the front upper part of the body on outer clothing. Pearson employees are required to sign a statement regarding proper badge usage when receiving a new or updated security badge. All facilities' entrances are monitored actively by security officers, receptionist staff, or closed circuit television systems.

Further access to restricted areas such as the Iowa City Data Center requires additional authorization, which is both programmed into the employee's badge and illustrated on the badge. Access is pre-approved on a business-need basis by the authorized manager.

Visitors may only enter Pearson facilities at designated entrances. Manager authorization is required for visitors, and they must remain with an escort (an authorized employee). Visitor badges must be worn in unobstructed view (same requirement for employees). Escorts are required to communicate visitor responsibilities to the visitor.

Monitoring

Closed-circuit TV cameras monitor all entrances, and uniformed guards regularly patrol the premises 24 hours a day.

Data Center Security

Pearson's core IT infrastructure resides in the Iowa City Data Center. Access to the Data Center is strictly controlled and managed. If an employee has a business need to access the Data Center, the employee's manager must complete a request for extended access. The employee's manager, the Pearson Administrative Services manager, and the Data Center manager must sign the completed form before access can be granted. To enter, an individual must pass through three access card readers with separate authorization at each level.

Physical and environmental protection controls in place include the following:

- Card key access for building and work area entrances
- 24-hour security officer coverage (note: coverage is for entire North Dodge facility and is not specific to the Data Center)
- Raised floor in Data Center to protect against water damage

- Dedicated, redundant cooling system with humidity control
- Emergency lighting in Data Center
- Fire extinguishers rated for electrical fires
- B/C rated fire extinguisher
- Smoke, water, and heat detectors
- Emergency power-off switch by exit door
- Surge suppressor
- Zoned dry pipe sprinkler system and chemical fire suppression system
- Uninterrupted power supply for all equipment
- Power strip/suppressors for peripherals
- Power strip/suppressors for computers

Printer Security Procedure

The integrity of any assessment program requires that we develop and adhere to stringent security protocols and processes for all stages of the test administration process. Pearson will clearly communicate the protocols and processes to all individuals and groups who have access to content or data related to the development, production, administration, scoring, or reporting for PARCC. Significant training will support protocols for security during testing, providing protection for secure materials, including print production.

Materials Production

Pearson maintains stringent security in designing, proofing, printing, and binding test materials. Compositor and printing vendors working with Pearson are required to maintain the following security measures:

- Electronic files, negatives, and plates must be kept secure by printing supervisors until they go to press.
- Electronic transfer of files is to be conducted via SFTP.
- Only authorized personnel shall be permitted access to test files, negatives, plates, or printed copies.
- All plates and negatives will be destroyed upon completion of the contract. Used plates will be placed into secure closed containers until they are released for proper disposal.
- Authorized staff will shred all press or bind make-ready waste material at the end of each day’s press run.
- Each production run will be made under the close direction of the appropriate pre-press, press, and bindery supervisor for the project.

- During the manufacturing process, work-in-progress and completed materials will be covered and controlled.

Secure Storage in Pearson Warehouses

Together, Pearson warehouse facilities in Iowa and Texas provide 303,000 square feet of climate-controlled warehouse space with secure access and professional security guards. Using an aisle-selective racking system to facilitate capacity, we are able to store more than 55,000 pallets of secure materials at any given time.

Our sophisticated inventory and warehouse management system provides end-to-end inventory and tracking of stored materials. Once materials arrive at the warehouse, Pearson staff use hand-held scanners to catalog materials directly into the system. The location of the test materials is likewise recorded as they are moved within the facility. This approach allows the system to track and regulate the movement of stored materials, and it provides Pearson staff with detailed pallet and inventory reports.

Our comprehensive approach includes measures for security, materials monitoring, and efficient space allocation. Together, these tools and processes allow us to maintain security, facilitate storage capacity, retrieve stored documents, and stage materials for disposal at the proper time.

Efficient Storage. We can identify where PARCC materials are stored and retrieve them if necessary. Our warehouses are air-conditioned and humidity controlled to provide maximum document protection.

Disposal of Secure Test Materials

Pearson moves all test materials through a single, secure disposal path to mitigate risk. Our local recycler will transport the materials to their secure recycling facility in a locked truck.

Our contracts with local recyclers define confidentiality clauses, including their enforcement. Our vendors have current security and alarm systems in place at their facilities to keep unauthorized persons from restricted work areas, and their employees sign confidentiality agreements. Pearson also conducts random audits to confirm adherence to our security provisions.

At the recycling facility, materials are securely destroyed, so that student data and test items and content are not compromised. As a final precaution, the material is shipped in sealed containers to a paper mill.

Cheating Detection

Caveon Test Security

Caveon provides test security products and services for testing and measurement. It is the first organization of its kind to provide an integrated solution for the prevention, detection, and remediation of test fraud (i.e., cheating, test question piracy, and proxy test taking).

Internet and social media monitoring scope of work will be performed by Pearson's subcontractor Caveon. Caveon will perform Web Patrolling services for the Partnership to evaluate threats to the administration and security of PARCC's summative assessments. Caveon will:

- Systematically patrol the Internet, websites, blogs, discussion forums, video archives, social media, document archives, braindumps, auction sites, and media outlets
- Identify and verify threats to PARCC test security and notify Pearson (who will notify PARCC as required)
- Once a threat is verified, work systematically through the steps necessary to have infringing content removed
- Provide summary reporting to including overall and specific threat analysis, with actionable recommendations for PARCC to follow in minimizing and removing the dangers and threats

Monitoring, Detecting, and Evaluating Possible Misconduct

In addition to Caveon's monitoring the Internet and social media, as discussed in **V.C. Psychometric Services**, Pearson will perform numerous data forensic analyses to investigate for statistical evidence of possible misconduct. As stated in that section, proposed statistical methodologies will include tracking excessive rates of response changes, evaluating person-fit, and identifying excessive levels of response similarity among pairs of test-takers.

In regards to the documentation requirement, Pearson will provide results files summarizing findings from data forensic analyses with flagged cases noted. Pearson recommends that PARCC representatives conduct follow-up investigations and gather additional evidence of a non-statistical nature confirming that misconduct has occurred prior to imposing punitive actions on test-takers or educators. Upon confirmation of misconduct, Pearson will work with PARCC to take necessary actions.

Documentation of Processes and Test Administrator IDs

Security Procedures for District and School Staff, Test Coordinators, and Test Administrators

District- and site-based personnel are a critical link in maintaining the security of PARCC materials. Security is everyone's concern and responsibility. We have developed a general overview of security issues, cautions, and instructions for PARCC states, district and school staff, including district coordinators, school coordinators, and test administrators, as described below.

Assessment Coordinators

Document security information will be included in test coordinator manuals for PARCC, in the directions for administration in the test administration training workshops, and on PARCC website. Handling and return procedures will detail the receipt, storage, administration, retrieval, and return of materials. The information will cover the following procedures:

- Security agreements for district coordinators and security affidavits for all district and school personnel who will have access to the testing materials. Pearson will work with PARCC to determine which state, district and school personnel are required to sign the online security forms. These forms may include – Reports of Security Breach, Irregular Testing Conditions, and Test Administrator Security Agreement.
- Security document checklist for district coordinators, with specific security warnings and instructions
- Receipt procedures for test materials to verify that all materials were received; instructions on steps to rectify material shortages before testing begins
- Procedures for securely storing test materials
- Procedures to prevent unauthorized persons from accessing test booklets
- Instructions for distributing site coordinator manuals and test booklets prior to testing dates
- Inventory procedures for handling test materials at each point in the testing process to maintain accountability and integrity
- Procedures for collecting, accounting for, and returning all test booklets and answer documents after testing

District and School Coordinators

District coordinators will be expected to take all necessary precautions to safeguard all PARCC tests and test materials by limiting the access of persons within the school district. The district coordinator will be required to sign a security agreement in which he or she agrees to be responsible for the following:

- Keeping on file the names of all persons having access to tests and test materials

- Requiring all persons having access to the materials to sign a security affidavit and keeping a file of the affidavits in the district office
- Keeping the test materials in a secure, locked storage area when they are not in use by students
- Monitoring and tracking test materials inventory and confirming that tests returned for processing or destruction are properly accounted for and paperwork is correctly filled out
- Providing secure transportation of test materials within the school district

District and school coordinators are also responsible for inventory control. They will use a test book security form and an inventory form to track and monitor test materials. The test book security forms for the school and district will be used to record booklets received, booklets returned, and any discrepancies in the test booklet sequence numbers.

The school-level test book security form will list the security numbers of the test booklets assigned to a school. The district-level test book security form will list the security number ranges of the test booklets assigned to the district.

District and school coordinators must use the appropriate test book security form to complete the following tasks:

- Inventory the booklets on receipt
- Report discrepancies within one working day of receiving materials
- Note any discrepancy or missing booklets
- Inventory booklets after testing
- Indicate, for any booklet not being returned, the booklet number and the reason the booklet is not being returned

In addition to district and school coordinators, test administrators also have a responsibility to maintain the security of all materials while they are in their possession. Manuals for test administrators will emphasize security.

Test Administrators

PARCC test administrators will receive test materials from their school coordinator. PARCC booklets are secure and must be returned to the school coordinator after testing. Each test booklet has a unique number and barcode printed on it. The school coordinator will keep a record of the booklet serial numbers provided to each test administrator and provide a record of any missing materials to Pearson after testing.

To maintain test security, PARCC test administrators must account for all assigned test booklets and answer sheets before, during, and after test administration. All PARCC test booklets must be properly locked and stored prior to and after administration. Test booklets must be collected and counted and any missing test booklets must be located prior to dismissing students from a testing session.

Additionally, for each test session – in paper and online testing, Pearson will collect test administrator ID's for security tracking purposes.

Requirement

V.B.2.C. Test Form Construction

1. Form Construction Goals
2. Blueprint Sets
3. Form Construction Process
4. Stage 1–Forms Construction Specifications and Requirements
5. Stage 2–Form Pulling
6. Stage 3–Forms Composition
7. Stage 4–Forms Review
8. Proposed Form Construction Review Schedule
9. Core Form Pulling Participants' Travel Costs–In-Person Meetings

Response Requirements for Section V.B.2.C.

b) Offeror's proposal shall include a response to all of the requirements specified in Section V.B.2.C.

c) In response to the requirements listed in Section V.B.2.C.1, the Offeror shall demonstrate an understanding of the Partnership's goals for the creation of forms. The Offeror may propose manual or automated form construction methodologies. One of PARCC's stated goals is for the ability of a student to take any test at any time without compromising security. This is certainly a goal within a pre-defined test window; however, PARCC's long-term vision is on demand assessments, using forms that meet PARCC's blueprint, available at any time a student is prepared to take the assessment. For pricing purposes, the Offeror should assume that test administration will occur during pre-defined assessment windows; however, the Offeror is asked to propose a path to achieve this long-term goal and to comment on the potential benefits, risks, and challenges that PARCC should consider.

d) Offerors shall include a description of the process they will use to identify a PARCC-wide representative field test sample for the ELA/Literacy field test described in Section V.B.2.C.2.

e) Offerors will propose comprehensive solutions for conducting virtual form pulling reviews and in- person form pulling reviews as referenced in Section V.B.2.C.5. Offerors may propose an alternate form pulling review process if significant schedule and/or cost benefits would result from a different approach.

f) Offerors will propose comprehensive solutions for conducting virtual form reviews as described in Section V.B.2.C.7.

g) Offerors may propose an alternate form review process if significant schedule and/or cost benefits would result from a different approach.

h) Offeror shall recommend the duration for each of the meetings listed in Table V.B.2.C.8.

i) If the Offeror determines that the number of planned meetings referenced in Table V.B.2.C.8. is insufficient to review the Partnership-determined number of test forms to be reviewed by these committees, the Offeror shall propose an alternate solution. Solutions may include supplemental virtual and/or in-person reviews within the development year, but the supplemental reviews must adhere to the meeting guidelines outlined in this RFP.

j) Offerors must include budgeted amounts in their Price Response/Reply and Budget Worksheets to cover the all reimbursable costs for all participants' air travel, car rental, lodging, meal per diem, and miscellaneous expenses such as parking, tolls, vicinity mileage, and other costs Offerors identify as necessary to participants' travel to the Core Form Reviews.

Deliverables for Section V.B.2.C.

a) Deliver Test Construction Specifications

b) Deliver Form Design Templates

c) Deliver Computer-Based Forms

d) Deliver Paper-Based Forms

e) Deliver FPR Meeting Plans

f) Deliver FPR Review Meeting Agendas, Participant Lists, Meeting Notes

g) Deliver FPR Review Training Materials

h) Deliver FR Meeting Plans

i) Deliver FR Review Meeting Agendas, Participant Lists, Meeting Notes j) Deliver FR Review Training Materials

k) Deliver Form Pulling Schedule

l) Deliver Form Review Schedule

Response

Pearson's content team with support from our psychometric team and with direction from the ETS psychometric team will be responsible for the test construction activities for the PARCC Operational Assessments program. This section will detail form construction goals, the number of blueprint sets to be produced, the form construction process which Pearson and ETS will follow in detail, the proposed form construction review schedule, and the discussion of the costs around in-person core form pulling meetings.

In the assembly of summative operational test forms for PARCC ELA/Literacy and Mathematics tests, Pearson will follow a process designed to meet PARCC's primary goals:

- Scores are interchangeable across forms within years and across years.
- Scales are established to both support the classification of students into one of five performance levels and maximize measurement information at the tails of the distribution.
- Time spent on items that do not count toward test taker's score is minimized.
- Security risks through over-exposure of items are minimized.
- The number of items/tasks available to be released each year is maximized.
- The number of parallel forms available for each PBA and EOY test is maximized.
- Cost is minimized.
- Computer-based form administration is flexible without compromising security.

PARCC Form Construction Design Goals

To meet these goals, Pearson will construct PBA and EOY operational forms for each content area (see below) and will use an optimal combination of operational items and a small number of matrix-sampled items on each form. The final operational forms will have internal linking items (both within and across years). The matrix-sampled items will include sets of items that link forms vertically (across grade levels) and within sets of field test items.

Pearson and ETS will develop detailed operational form construction specifications (see below) for PARCC's review and approval. Given these specifications, we will evaluate the pools of field-tested items for ELA/literacy and mathematics to begin the process of selecting and allocating items and tasks for use on the summative forms. Our primary guide in selecting and allocating items will be the PARCC blueprints for ELA/literacy and mathematics, which display the following for each grade level and content area:

- Types of items and tasks
- Number of points associated with each item and task type
- Number and kinds of stimuli associated with each item and task type
- Distribution of items and tasks (PBA or EOY) (and associated points) across content domains/strands
- The total number of items and points, for each component as well as for the combined test

Operational forms will be developed to meet well-specified statistical targets to meet the needs of the PARCC assessment system, developed to industry standards for validity, reliability, and fairness (Standards for Educational and Psychological Testing, 1999) as informed by the results of the field test. The role of these targets is described in Stage 1 below. Historical experience in K–12 test construction, where resources for item development are not infinite, indicates that some trade-offs of test construction goals may be needed for some test forms. Pearson and ETS will work closely with PARCC to develop clear criteria for such trade-offs, if they are required.

An effective forms assembly process will provide measurement along the full range of ability, resulting in scales for ELA/literacy and mathematics that enables classifying students into one of five performance levels. To enhance the reliability of student classification based on this newly established scale, Pearson content specialists and psychometricians will work together closely to determine and evaluate the distribution of items across all ability levels when selecting items to establish, confirm, or maintain vertical scales, as applicable.

During the forms assembly process, we will also analyze and evaluate information obtained at the very low and very high ends of the scale (minimize CSEMs, etc.) in order to maximize measurement information, while maintaining other content and psychometric goals. ETS psychometricians will then do an independent, final review of the statistical properties of each form, as well as adherence to specifications for horizontal and vertical linking.

The final forms will include a core form with a set of matrix items. With the variety of mathematics items and points per item, it is not feasible to have the same number of items per matrix section; however Pearson recommends that the student experience be as parallel as possible. In determining the number of points for mathematics and the items for ELA/literacy, we worked to have students take approximately the same number of items, minimize the testing time, and reduce costs for additional forms.

Pearson is proposing one additional text set be added to the ELA/literacy EOY. Each text set would be composed of the passage set and six items. Most texts would appear on two forms with a different mix of items. For mathematics PBA, we are proposing a matrix section totaling six points and for EOY one with eight points (3–4 items). This will add 15–20 minutes for each embedded field test section, using 2.5 minutes per point as a guideline.

Determining when and how to field test ELA/literacy PBA is challenging. Embedding another task set would add another 45–60 minutes to the PBA which is challenging for the student and adds more costs. PARCC has proposed administering a PBA task set during the EOY testing window as a standalone field test, which we agree is the best method. An ELA/literacy field test form will consist of one task set with 6 to 10 items, one of which will be a Prose Constructed Response (PCR). All task sets of a particular kind will have the same number of items, so the student and classroom experience will be parallel. We have planned more forms in year 1 than the others due to having eight blueprints of field test items available.

With any high stakes testing effort, there is a security risk. With the longer testing windows and a larger population of test takers, the risks increase. The risks can be lessened by using multiple test forms that are or can be equated. Test forms can be administered in a thoughtful way, such as exposing a small number of forms each week, monitoring the number of students who have taken a test and having a scheduled or rule-based administration policy moving online forms in and out of the administration window.

The release plan is a PARCC policy decision. Pearson supports the goal of releasing test items for instructional purposes. Items or forms can be selected based on exposure. Another methodology for releasing items is to select from model items representing skills students are struggling with and whose statistics are such that they will likely not appear on a test form. Pearson will work with PARCC to determine and implement the release policy as discussed in V.B.5 below.

One strategy PARCC plans to use to minimize the cost of item development and reduce the exposure of items/forms is to build multiple parallel forms from a smaller pool of items, where items can and will appear on multiple forms. This could be accomplished by building full forms or possibly by developing modules/testlets to specification such that the form is built of a set of modules or testlets.

PARCC desires flexibility in the administration of computer-based forms. Computer-based tests do not require printing and shipping, so they are easier to administer during a wider administration window. While having forms available 24/7 is possible, administration constraints and policies will be necessary. We will work with PARCC to determine who will be taking these forms, the level of monitoring needed, how often a student can test, and how to mark forms so students do not get the same form or ones that have a significant number of common items. Security of test forms, as well as confidence in students' scores, is a definite concern. Other considerations, including the cost or turn time for scoring, may increase depending on the volume.

Blueprint Sets

To meet the assessment administration goals, Pearson will use an optimal combination of operational items to build a set of 10 core online forms and 6 core paper forms with a small number of matrix items on each EOY and mathematics PBA form. For the first year, we will use the anticipated five blueprint sets provided under another PARCC contract to build the forms. Items will be used on multiple forms to yield the desired 10 core online forms. In the following years, the form counts will be as described above except for algebra I, geometry, and integrated mathematics will change to 12, 12, and 8, respectively, and the number of core paper forms is reduced to 4 in year 4. The items used on the core forms in years 2-4 will be a combination of items from the last field test and active items previous years.

In the first two years of the contract, Pearson anticipates online field testing the item distribution shown in following tables for mathematics and ELA/literacy. For 2014-15, these items will be developed under another contract. For 2015-16, Pearson will develop approximately 80% of the field test the items with the rest coming from another PARCC contract. These counts were developed based on current blueprints and attrition rates at data review provided by PARCC. The mutually agreed upon development plan may shift the numbers somewhat.

	Performance Based Component						End of Year Assessment	
	Type I Items		Type II Items		Type III Items		Type I Items	
Grade / course	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items
3	63	50	30	20	23	15	245	195
4	63	50	30	20	23	15	225	180
5	57	45	30	20	23	15	200	160
6	63	50	30	20	23	15	215	170
7	63	50	30	20	23	15	208	165
8	70	55	30	20	23	15	210	165
Alg 1	75	60	36	24	36	24	268	210
Geo	75	60	36	24	36	24	260	204
alg II	63	50	38	25	38	25	217	170
Int I	50	40	24	16	24	16	173	136
Int II	50	40	24	16	24	16	173	136
Int III	50	40	30	20	30	20	177	140

	Performance Based Component								
	EBSR		TECR		PCR		Passages*		
Grade	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Passages	Total Field Test Item Count
3	154	60	56	25	42	15	35	25	252
4	175	75	56	25	42	15	42	30	273
5	175	75	56	25	42	15	42	30	273
6	175	75	56	25	42	15	42	30	273
7	175	75	56	25	42	15	42	30	273
8	175	75	56	25	42	15	42	30	279
9	200	90	64	30	48	18	48	36	312
10	200	90	64	30	48	18	48	36	306
11	175	75	56	25	42	15	42	30	231

	End of Year Assessment								
	EBSR		TECR		PCR		Passages*		
Grade	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Passages	Total Field Test Item Count
3	192	90	64	40	0	0	32	25	256
4	192	90	64	40	0	0	32	25	256
5	192	90	64	40	0	0	32	25	256
6	192	95	64	35	0	0	49	30	256
7	191	95	67	35	0	0	49	30	258
8	191	95	67	35	0	0	49	30	258
9	222	114	78	42	0	0	57	34	300
10	222	114	78	42	0	0	57	34	300
11	191	95	67	35	0	0	49	30	258

The following tables provide the planned field test counts for the online assessment for the last two years of the contract, based on current blueprints and attrition rates provided by PARCC. The mutually agreed upon development plan may shift the numbers somewhat.

	Performance Based Component						End of Year Assessment	
	Type I Items		Type II Items		Type III Items		Type I Items	
Grade/ course	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items
3	51	40	24	16	19	12	196	156
4	51	40	24	16	19	12	180	144
5	46	36	24	16	19	12	160	128
6	51	40	24	16	19	12	172	136
7	51	40	24	16	19	12	167	132
8	56	44	24	16	19	12	168	132
Alg 1	63	50	30	20	30	20	224	175
Geo	63	50	30	20	30	20	217	170
alg II	51	40	31	20	31	20	174	136
Int I	25	20	12	8	12	8	87	68
Int II	25	20	12	8	12	8	87	68
Int III	25	20	15	10	15	10	89	70

	Performance Based Component								
	EBSR		TECR		PCR		Passages*		
Grade	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Passages*	Total Field Test Item Count
3	132	52	48	22	36	13	30	20	216
4	150	65	48	22	36	13	36	24	234
5	150	65	48	22	36	13	36	24	234
6	150	65	48	22	36	13	36	24	234
7	150	65	48	22	36	13	36	24	234
8	150	65	48	22	36	13	36	24	240
9	175	79	56	27	42	16	42	30	273
10	175	79	56	27	42	16	42	30	267
11	150	65	48	22	36	13	36	24	534

	End of Year Assessment								
	EBSR		TECR		PCR		Passages*		
Grade	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Items	Total Items Field Tested	Projected Operational Passages	Total Field Test Item Count
3	165	78	55	35	0	0	30	20	220
4	165	78	55	35	0	0	30	20	220
5	165	78	55	35	0	0	30	20	220
6	165	82	55	30	0	0	36	24	220
7	164	82	58	30	0	0	36	24	222
8	164	82	58	30	0	0	36	24	222
9	195	100	69	37	0	0	42	30	264
10	195	100	69	37	0	0	42	30	264
11	164	82	58	30	0	0	36	24	222
	* Note: These counts represent single passages. Some item sets require multiple passages.								

In the previous section, the plan for field testing new items was discussed. As specified in the RFP, each year, we will identify a PARCC-wide representative sample sufficient to gather and score 1,500 valid student responses per field-tested task. Field test forms will be spiraled at the student level to obtain the most representative sample possible for each item. Student demographics and the characteristics of the states which comprise the PARCC consortium are relatively diverse.

Therefore, to verify adequate representation, post-stratification will be used to identify a targeted sample of 1,500 valid student responses for hand scoring each embedded field test item. We propose oversampling by approximately 20 percent or 300 students per task in order to obtain 1,500 valid responses per task. In some cases it may be necessary to oversample even further in order to obtain 1,500 valid responses, particularly for items that are very difficult or very easy.

Proper execution of a robust sampling plan forms a solid foundation for analysis. We understand the critical impact representative sampling has on calibration, scaling and equating, and bring that understanding to our work on the PARCC Summative Assessments. ETS is currently providing sampling plans for PARCC field testing and Smarter Balanced, and we have done so for a number of K12 clients (e.g., Washington, Virginia, New Jersey, and California). We will work with PARCC and your TAC to verify that the sampling plans for hand scored items will yield the most accurate results possible, in the most efficient way possible.

Form Construction Process

The process for form construction is multi-step and very iterative. Many people are engaged, including content and psychometricians from the contractors and PARCC, publishing staff, editorial, and proofreading. PARCC has proposed a four-stage process for form construction. Pearson will follow that process as outlined in the RFP.

Stage 1: Forms Construction Specifications and Requirements

The operational form construction process will begin with a collaborative effort between PARCC, Pearson, and ETS to develop comprehensive test specifications for each content area and grade/course, for computer- and paper-based forms, and for braille, large-print, and text-to-speech forms.

Pearson content specialists and ETS psychometricians will employ a research-based approach to the development of the test form specifications and target statistics, which will reflect best practices in the field. The specifications will provide a detailed rationale for the proposed number of forms to be developed, as well as the form design templates including, at a minimum, the number of tasks, the number of linking items, and the degree of item overlap. These specifications will be reviewed and approved by PARCC prior to implementation.

Pearson and ETS will provide well-defined test construction specifications to support production of forms that are parallel with respect to content, construct representation, and statistical properties. The extent to which this is realized will depend on the depth and breadth of the available item pool. The specifications will describe how to build operational forms to be comparable (within content and grade/course) in terms of test form difficulty and discrimination.

The specifications will describe the rationale for choosing psychometric properties such as test characteristic curves (TCCs) and conditional standard errors of measurement (CSEMs), as well as the manner by which the operational test forms will be constructed to match these targets. We will evaluate TCCs, test information function (TIF) curves, and CSEM curves for each operational test form, examine performance attributes captured by PARCC's performance level descriptors. The TCCs and TIF and CSEM curves and summary statistics for each operational test form will be provided to PARCC as part of the forms-construction documentation.

All other factors being equal, administration of truly parallel forms would eliminate the need for statistical adjustments of scores from one form to another. However, without an infinite item pool with perfectly accurate item statistics, true parallelism of multiple test forms is an unachievable ideal. As a result, statistical adjustments to account for differences in form difficulty are implemented to support comparability of scores from one test form to another. Methods to implement these adjustments may include equating, linking, or concordance. Depending on the situation, adjustments may take place during the test construction process (pre-equating) or following administration prior to release of scores (post-equating).

Form construction specifications must address the program-specific linking strategies to adjust for differences in student scores that may occur due to differences in test forms. If linking is achieved through use of common items, specifications for the selection and placement of linking items include content and statistical considerations that reflect the criteria for the operational forms.

Use of test forms that are constructed to be parallel with respect to content and psychometric properties, along with statistical procedures (equating, linking, concordance) to adjust for differences in form difficulty will support comparability of student-level scores in accordance with the Standards for Educational and Psychological Testing (1999). As applicable by content and form, the specifications will provide details as to the strategy for selection and placement of linking items and will document the distribution of content for alignment to the blueprints for ELA/Literacy and Mathematics.

Form construction specifications will include criteria for selection of matrix-sampled items to support the need of comparable scores, establishment and maintenance of vertical scales, and replenishment of the item bank through field testing. Criteria for selection of matrix-sampled items for individual forms will take into account the appropriateness of the mix of content and item types relative to the set of operational items, and the impact on the testing experience of the individual test taker.

Strategies for distribution of matrix sampled items across forms, whether for linking (horizontal or vertical) or for field-testing, and distribution across students will take into account the number of items to be matrix-sampled as well as the need to minimize individual testing time and minimize security risks through item over-exposure.

Stage 2: Form Pulling

Form construction for the first two years of the contract relies on receiving inputs from other PARCC contracts. These inputs include field test data, results from data review, scoring materials, up-to-date permission information, and identification of items that are field test ready from the PARCC Administration contract. It also includes being provided the form pulls for Fall Winter Block 2014 forms. Pearson will coordinate tasks with the current PARCC item development team, PARCC Field Test administration team, and PARCC States and Partnership manager. . Items and item-level data are expected to be available through the item bank. We suggest that a planning meeting be set up to discuss process, procedures, and schedule to facilitate a smooth transition.

Pearson will use an automated process to make the initial pull of core test forms. Item statistics, item metadata, psychometric targets, and parameters from the form constructions specifications will be used in this process. Once the operational-ready item pool has been populated with the most recent field test data and the parameters are programmed into the test assembly tool, then the form pulls will be initiated, using the item pool available for forms construction. ,. Pearson will consult with PARCC about specific models for the operational form assembly. Since it is PARCC's plan to use items, scenarios, and text sets across forms, one model may be to build parallel partial forms or modules.

Forms will be selected for online and paper core forms first. After the initial pull, content specialists will review the forms, checking for clueing, variety, and adherence with the form construction specifications. Psychometricians will review the item characteristics of the assembled test form to establish that the items are appropriate in terms of test form equating plan and statistical specifications.

Content specialists and psychometricians will work together to revise the form as needed. Each core form will be evaluated using series of assembly-engine generated reports and a checklist based on the form construction specifications.

The ultimate approval of the forms resides with PARCC. Once our content and psychometric staff have released the forms for review by PARCC, an in-person meeting will be held at a Pearson's site. Committees arranged by grade band will review the 10 online and six paper forms per grade/course for year 1   In the following years, the online form counts for Algebra I, Geometry, and Integrated Mathematics (I, II, and III) will change to 12, 12, and 8, respectively, based on the planned field test counts. The paper form count will reduce to 4 in year 4 and earlier for Integrated Mathematics. PARCC reviewers will have access to the items and item list for each form. For ELA/literacy, complete sets will be available to facilitate the review of the form. The reports and checklist generated during test assembly will be available to PARCC.

PARCC representatives can evaluate and make decisions related to areas of concern, if any exist, for a particular form. It is anticipated that all forms will meet the blueprint, but it may be challenging to meet all of the psychometric targets. We intend for each test form for each content and grade/course area to meet all psychometric targets, and will work with the Partnership representatives and your technical advisors until you are satisfied with the statistical properties of the forms, as well as with blueprint coverage.

Matrix items will be embedded into the approved Core forms for mathematics PBA and EOY for both ELA/literacy and math. These embedded sections will be populated with vertical linking items and newly developed items that PARCC has approved. Due to the variety of item types and point values of items in mathematics, the embedded sections will have some variability. There will be multiple configurations to the field test section that impact answer documents for paper forms. Consumable test booklets will be used for grade 3 for both mathematics and ELA/literacy for both PBA and EOY administrations. We will provide the following number of unique answer documents each year for ELA/literacy: Operational PBA (3 per grade level 4-11), EOY (1 per grade 4-11), and PBA Field Test (3 per grade level 4-11). For mathematics, items in the field test sections can be configured such that no more than 60% unique answer documents will need to be developed each year. We will include explicit instructions regarding the Answer Documents in the Ancillary Assessment Materials.

Pearson will work with PARCC to reduce the impact of this variability, which extends to the classroom where administrators may have to manage multiple versions of answer documents. As described earlier in this section, we are proposing a couple more points be added to the matrix section for EOY and PBA, adding approximately five minutes to the testing time. Pearson will consult with PARCC on the configuration of the matrix sections. Content specialists will check the matrix form for cluing with items on the core form in which it is targeted to be associated with, and combine it with another form if needed.

After the sets have all been assigned to a core form, PARCC will review the forms. To make more efficient use of reviewers' time, all matrix sets to a respective core form will be grouped together for the review. ELA/literacy representatives will also review the PBA field test forms. Text sets will be reviewed together.

PARCC representatives will use a checklist for the Form Pulling Review. We will work with PARCC to finalize the checklist and develop training materials for the review of the core forms and the full forms that include the matrix items. The review of the core forms will be in-person, arranged by grade band. At least two representatives will review each core form. The accessibility expert will review all forms targeted for accessibility and accommodations for adherence to the Accessibility Guidelines.

We expect that the selection and approval of accommodated form for closed-caption, ASL videos, braille, large-print, and text-to-speech for the various inclusion orders, including non-visual will be identified and approved by an accessibility representative in attendance at the meeting. The review of the full forms with the matrix items is a forms review, not a content review. Pearson assumes that we will receive approved/clean items from the existing Item Development Contractors. As a result, we have not included time or costs for content edits, other than to correct a small number of mutually agreed upon fatal flaws.

Stage 3: Forms Composition

Once items have been selected and ordered, forms can be composed. Pearson will use items from the PARCC Item Bank to design and assemble both computer-based and paper-based test forms for PARCC's summative assessment. The PARCC Style Guide that includes the Editorial Style Guide and the Computer-Based Style Guide will be used as appropriately to compose the forms. We will work with PARCC to develop and document additional specifications as needed.

PARCC plans to provide three formats to access visually impaired students—braille, large-print, and text-to-speech. For the paper-based test, we will create electronic files in the format needed to produce one form per grade per assessment of braille and large-print. For refreshable braille and text-to-speech forms, we will provide those computer-based files. Pearson will have the materials proofed by an independent party who is a certified braille reader. We understand that PARCC may choose to employ the services of a braille proofreader at their expense. In addition, visual experts from PARCC states will be included in final review of forms.

Pearson will contract with a PARCC-approved publisher of braille and large-print materials. Answer documents, test administrator notes, and scripts to accompany the braille test versions will be developed and published.

Pearson will publish the computer-based forms within TestNav 8 delivery system. Planning discussions with the content providers under the Item Development Contract and Pearson need to happen during the first two years so that the items received under the Item Development and this contract are developed in a consistent manner with respect to QTI and APIP interoperability standards and no additional transformation work is required.

Pearson employs a proofing process that includes quality assurance and other proofreading checks by Pearson staff. It should also be noted that the PARCC items undergo rigorous reviews at every phase of the item development process and there should be no item level edits for operational items at this phase in the process.

Publishing Quality Assurance Checklist

- Confirm content renders as supplied
- Publishing styles applied consistently
- Technology enhanced items render and function correctly
- Customer deliverables are complete and accurate
- Form level tools operate correctly
- Resources (exhibits, directions, sample items, etc.) display correctly
- Navigation of the form is accurate (i.e. sections appear as required (including seal codes, items are in correct order, test exit options are correct, etc.)
- Media settings work as intended (includes text-to-speech)

Stage 4: Forms Review

Computer-Based Forms

The last step prior to administration is to have PARCC review the composed forms. The processes for paper and online forms differ slightly. This review includes all forms, including those for computer-based accommodations. The review is iterative with three rounds.

All computer-based forms will be reviewed during Round 1 (Alpha). Forms that contain errors will be corrected by Pearson. During Round 2 (Beta), those forms not approved in Round 1 will be presented again for PARCC to confirm that the edits have been applied correctly. During Round 3, a mock data review will be performed as discussed in Section V.B.2.H.

Pearson will develop a detailed schedule for managing the multiple review steps and the volume of the forms. Pearson will negotiate with PARCC and work to find a plan and schedule to review the large volume of forms through this process.

The forms will comply with the PARCC Style Guides and Pearson will work with PARCC to develop any additional specifications that are needed for this process. It will be essential that everyone is working from the same set of requirements and interpretations.

Forms Review Committee

Pearson expects that the membership on the Form Review Committee will remain constant from form pull reviews through mock data review. Checklists and processes will clearly define the committee member's work at each stage of the review process. We will work with PARCC to define the process and the mechanism for transmitting feedback and approvals.

All reviews after the in-person forms pull meeting will be virtual. Committee members will review items in a secure system. All materials presented to the Committee members will be reviewed internally prior to the forms review. No stipends are planned for this work.

Paper-Based Forms

These forms will also have three rounds of review. The criteria used in the reviews are very similar to those for the computer-based forms. It is expected that there will be some changes in the checklist to conform to the presentation requirements of paper. The last round is a digital proof round, which is described in the **Requirements for Print Materials** section.

Proposed Form Construction Review Schedule

Pearson will work with PARCC to determine when data will be available for the construction of the operational forms. A draft schedule has been included in the Other Supporting Material section of this proposal.

Core Form Pulling Participants' Travel Costs—In-Person Meetings

As described in Section V.A.1.G, the cost of participants' air travel and/or rental car (as needed), lodging, meal per diem, and miscellaneous expenses such as parking, tolls, mileage, and other costs are included in our Cost Proposal. Additionally, Pearson will be responsible for meeting room and equipment costs.

At least eight weeks prior to the review meeting, we will provide a meeting plan which will include the meeting agenda, participant list (to be updated prior to the meeting with final list of educators who plan to attend), and other required information. Below is a list of the Forms Construction Meetings that will be held each year. The first Data Review is in July 2015, following the first operational administration, which includes a number of embedded field test items.

Meeting	Year 1 Number of Virtual Meetings	Year 1 Number of Virtual Meetings	Meeting Duration Per Meeting (Days)	# of State Participants Per Meeting	# of PARCC Staff Attending Per Meeting	# of PARCC Staff Attending Per Meeting	Location
Common Form Pulling	2	0	3	20	4	14	Contractor Site
Matrix Form Pulling Review	0	2	5	18	3	12	Virtual
Data Review	1	0	5	118	12	19	Hub City

Requirement

V.B.2.D. Training Materials and Test Manuals

1. Training Materials
2. Test Coordinator and Test Administrator Manuals

Response Requirements for Section V.B.2.D.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.D.

Response

Training Materials

The implementation of the new PARCC Operational Assessments will require some changes to processes and procedures for states, districts, and schools compared to how their current state assessments are administered. Effective training is essential for smooth administration of paper-based and online testing. Our training materials will focus on all facets of the new administration, from using the PARCC Administrative Portal for registration and enrollment, how to order materials, the receipt of paper-based materials in the district, the distribution of materials within the district, collecting materials for return to Pearson, and the policies and procedures for maintaining security at all times. We will also include training for LEA and School Technology Coordinators for online administrations and security procedures related to online testing.

Detailed training materials will be provided to all LEA participants for PARCC administrations, including LEA test coordinators, school test coordinators, test administrators/proctors, LEA technology coordinators, school technology coordinators, and any others who may need to participate in setup or administration of PARCC assessments (such as LEA or school personnel involved in special needs populations requiring accommodations or accessibility technology).

Training materials will cover topics including student registration procedures administration protocols, security policies/protocols/procedures, the technology delivery system, and accessibility and accommodations policies and protocols.

Additionally, training materials will address both paper-based and computer-based administrations, and will be modularly structured by topic, intended audience, and with alternate versions of materials to consider the varying experience levels of LEA and school staff. Pearson will provide 14 different types of modular trainings.

The training materials will be provided as a Microsoft® PowerPoint presentation or in another PARCC-approved format with up to 100 slides for Windows® and Macintosh® platforms for each training, along with explanations accompanied by up to 50 full-color graphics depicting relevant items.

Training will be delivered via narrated PowerPoint web presentations using pre-recorded WebEx or other similar mode and made available on the training site. Electronic copies of all training materials will be available so states can customize the training materials on their own for training purposes. Schools can print the materials locally, as needed; Pearson has not included costs for hard copy modular training materials.

A proposed structure for these modular trainings is summarized in the figure below. The range of topics will comprehensively cover the aspects of preparing for and administering both paper-based and computer-based assessments, as well as the delivery and interpretation of reports. Some topics may be intended primarily for testing coordinators (with sectional organization to address LEA and school-building-specific information needs), others may be intended primarily for technology coordinators, and some may be equally appropriate for both.

Additionally, each fundamental training topic will be available in two separate varieties. Training modules “For Staff New to PARCC Assessments” will be comprehensively organized to provide training for anyone who may be new to PARCC assessments or new to online or paper-based assessments in general. These training modules will provide all basic information, assuming little to no prior knowledge on the part of trainee audiences.

Training modules “For Staff Having Prior Experience with PARCC Assessments” will cover the same general topic categories, but for audiences who Pearson assumes will have already made use of the more comprehensive “For New” training materials in previous years.

Training material for staff with prior PARCC assessment experience will instead focus on both essential “refresher” content for successful administration of the assessments, as well as in-depth highlights on any policies, protocols, or procedures that may have changed since the previous year—including changes driven either by technology (introduction of new platforms or operating systems, sunsetting of outdated platforms, or similar technology-related updates) or changes related to recent policy decisions.

Where appropriate, the topical and modular approach for staff with prior experience may represent a combining of topics otherwise kept separate for staff who are new to the assessment training material. This is illustrated in the proposed modular approach in the following figure.

Proposed Modular Approach for Training Materials			
Intended Audiences		**Topical Organization**	
Testing Coordinators	**Technology Coordinators**	**For Staff New to PARCC Assessments**	**For Staff Having Prior Experience with PARCC Assessments**
✓		Administration of Paper-based Assessments	Administration of Paper-Based and Online Assessments
✓		Administration of Online Assessments	
✓		Obtaining and Interpreting Reports	Obtaining and Interpreting Reports
✓	✓	Technology “Readiness” for Schools & Districts	Readiness for PARCC Assessments
✓		Using Practice Tests for “Student Readiness”	
✓	✓	Infrastructure Trials: Running a Dress Rehearsal	
	✓	Proctor Caching & Network Data Mgmt.	Proctor Caching and Network Data Mgmt.
	✓	Test Security Policies & Procedures	Test Security Policies & Procedures
✓	✓	Accommodations & Accessibility	Accommodations and Accessibility

PARCC will review and approve the overall modular and topical organization of training materials, as well as all individual training modules, prior to their release. All materials will be proofread by Pearson and also provided to PARCC representatives for proofreading, prior to being finalized and/or printed. The modules will be made available on both PARCC and Pearson websites 60 days prior to test administration.

Test Coordinator and Test Administrator Manuals:

Pearson will develop Test Coordinator and Test Administration Manuals for the operational assessment and collaborate with other PARCC contractors to fully address PARCC’s needs. Each manual will have a unique design to easily distinguish between Test Coordinator and Test Administration Manuals and computer versus paper-based testing and grade bands.

All manuals will be published in print, and will be available through the Pearson and PARCC websites in an ADA compliant format. Manuals will be thoroughly proofread and in their final state before seeking PARCC’s approval to print or publish on the websites. All manuals will be published online and paper copies will arrive in districts and schools 60 days prior to the PBA test administration and will be available online throughout all administrations. For more information, see the Materials List (revised 4/6/14).

The Test Coordinator Manual will cover the receipt, distribution, collection, and return of materials to Pearson, and the training of test administrators and detailed instructions on security policies and protocols approved by PARCC. These manuals will align to the training provided in the modules detailed in section one above. The manuals will provide comprehensive directions on how to administer the operational assessment. We will develop separate Test Administration Manuals for the computer-based and paper-based administrations. We will incorporate lessons learned and any improvements from the field test administration into the operational manuals.

The training materials for the Test Administrator Manuals will be updated from the field test administration to reflect any identified areas for improvements. The manuals will include requirements and scripts for all assessed grade levels and subjects. Instructions for accommodated testing (e.g., braille, text-to-speech, and others) will be provided in addition to the manuals. The computer-based manuals will cover the practice activities and scripts to familiarize students with the computer-based test.

The following table summarizes the Test Administrator Manuals that will be provided. In addition, districts will receive a Test Coordinator Manual for each mode of administration, as needed, for computer-based testing and paper-based testing which covers topics across grade bands and subjects.

PARCC Test Administrator Manuals Combined ELA/Mathematics Manuals for Grades 3-5, 6-8		
Grades 3-5	**Grades 6-8**	**Grade 9-11**
ELA/Literacy and Mathematics for computer-based administration	ELA/Literacy and Mathematics for computer based administration	ELA/Literacy for computer-based administration
ELA/Literacy and Mathematics for paper-based administration	ELA/Literacy and Mathematics for paper- based administration	ELA/Literacy for paper-based administration
		One manual per course for computer-based administration: Algebra I Geometry Algebra II Integrated Math I Integrated Math II Integrated Math III

PARCC Test Administrator Manuals Combined ELA/Mathematics Manuals for Grades 3-5, 6-8		
Grades 3-5	**Grades 6-8**	**Grade 9-11**
		One manual per course for paper-based administration: Algebra I Geometry Algebra II Integrated Math I Integrated Math II Integrated Math III

The following figure summarizes the number of manuals required, and the length of each manual in terms of number of pages. As instructed in the RFP these counts were used for cost estimating purposes (plus 22 percent to the size of a single content area manual to accommodate the second content area specific instructions), however from our current field test experience we know that the number of pages is much higher. We will work with PARCC to keep the number of pages within this range, as opposed to adding to the cost of the program. It may also be possible to move some of the information into other training sources such as the PowerPoint, if it is more background information and less instructional.

PBA & EOY Manuals			
Ancillary Item	**Computer-Based**	**Paper-Based**	**Pages**
Test Coordinator Manual	1 per every LEA school	1 per every LEA school	96
Test Administrator Manual	1 per every 25 students	1 per every 25 students	140-184

For the combined Test Administrator Manual for grades 3-5 and 6-8, we have estimated adding 22 percent to the size of a single content area manual to accommodate the second content area specific instructions. This increase is more than offset by eliminating the second content area manual for the grade band.

The manuals will be distributed prior to the Performance Based Assessment testing window. Instructions will be included to retain the manuals for the End of Year Assessment. Given the size of the manuals and the number that need to be distributed (over 640,000 for the spring testing window), it is not efficient or cost effective to print and distribute separate (yet identical) manuals for the PBA and EOY testing windows.

The instructions will include a URL for districts to access replacement manuals online, which may be printed locally if the original hard copy manuals are misplaced. New manuals will be provided for each assessment window (Fall/Winter Block, Spring Block, and Spring Traditional) that the school or district participates in annually.

Pearson will update these materials each year based on lessons learned from previous administrations. We will work with PARCC to assemble a comprehensive, effective training program for test administrators in year one of the contract, and we will modify it as needed in subsequent years.

Requirement

V.B.2.E. Student Registration

Response Requirements for Section V.B.2.E.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.E.

Deliverables for Section V.B.2.E.

a) Enrollment Report by State

Response

Student Registration

The PARCC Summative Assessment enrollment information for both the paper-based and computer-based administrations will be collected through Pearson's PearsonAccess system, where users will place orders for all testing and ancillary materials, including specialized editions like braille and large print tests.

Pearson will seek PARCC approval prior to releasing authorization to the system. Authorized users will use the system to enter registration counts for both paper- and computer-based tests. Accurate registration information from the districts enhances the ability to distribute appropriate quantities of test materials, and it reduces the amount of extra materials shipped. Because registration counts are the basis for the quantity of test materials ordered, authorized users will be able to make changes to order quantities or make material requests after the initial order has been placed. Starting in Year 2, we will preload previous registration counts into the portal so comparisons can be made.

One challenge that PARCC will likely face during the first operational administration is getting districts to make timely enrollment submissions. To assist with registration communications and reminders, Pearson will use the administrative system notification feature for sending emails to selected schools and districts. Registration announcements and reminder emails can be configured based on PARCC requirements, enrollment status, and contact responsibility.

Following the close of the registration window, Pearson will provide a summary of registration counts for each state (and district/school within the state) for both paper and computer-based tests to state and district test coordinators to review, amend as needed, and approve. As each registration window closes, a notification will be sent for each state to review and provide approval of their registration counts. Pearson will contact states to confirm registration counts for any states' enrollments still in review status.

The registration window will include a deadline for entering preliminary and final enrollment data. Enrollments for both PBA and EOY will be entered by districts at the same time. Below is a draft schedule. In order to finalize print quantities and complete resource planning for an administration, the preliminary counts should be within 10 percent of the final registration numbers, including anticipated split between paper and computer-based testers, which need to be finalized by the end of the registration window. Students at grades 3-8 should be registered for a single mode of administration (paper or computer-based) for both ELA and mathematics, unless the state has worked with Pearson to implement a specific exception to this requirement. At high school a student should be registered for a single mode of administration by test (e.g., Algebra I)—both the PBA and EOY component of the test should be taken in the same mode. If a high school student is taking multiple tests (e.g., Algebra I and ELA I) in the same test administration window he/she can take one content area online and the other on paper because the high school tests are priced by test not by student as they are done at the lower grades.

Administrative Window	Year 1 Approximate Timing of Test Administration	Year 1 Preliminary Registration Window	Year 1 Final Registration Window
Fall/Winter Block*	PBA: November-December EOY: December-January	August 4–10	August 11–September 5
Traditional Year	PBA: Mid-Feb-Mid-April EOY: Early April-Late May	October 6–26	October 27–November 14
Spring Block	PBA: Early April-Mid-May EOY: Early May-Mid June	December 15–26	December 27–January16

*For the Fall/Winter Block 2014 the PBA administration will be 12/1-12/19 and the EOY administration will be from 12/15-1/16. The Braille forms will be available the last week of the PBA and EOY administration. All testing for the Fall/Winter Block 2014 will be on paper only and for all high school tests except for the three integrated courses. The registration window is 9/1-10/1. Print quantities will be finalized based on counts as of 10/1. States can require districts to have counts entered into PearsonAccess prior to 10/1 as a preliminary window and to provide states with sufficient time to validate counts prior to the final enrollment. Pearson will work with PARCC for subsequent administrations to determine whether the entire registration window needs to be available to districts and schools, or only a subset.

Requirement

V.B.2.F. Requirements for Print Materials

Response Requirements for Section V.B.2.F.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.F

Deliverables for Section V.B.2.F.

a) Deliver Print Specifications
b) Deliver Test Books and Answer Documents
c) Deliver Mathematics Reference Sheets
d) Deliver Manipulatives

Response

Requirements for Print Materials

Pearson will provide high quality materials in a timely and efficient manner to meet PARCC's test administration needs. Since the majority of printing work is done in-house we control the production environment, press schedule, and quality process for print materials. We also employ strict security requirements to protect secure materials production. We will produce materials according to the PARCC Style Guide and to the detailed the specifications supplied in the Materials List (revised 4/6/14). .

Pearson understands PARCC's staffing constraints and will work with PARCC to establish an agreed upon form review schedule that will balance the time constraints in the Operational Administration schedule while maximizing the time allotted for PARCC reviews. Pearson recognizes the inherent challenges faced in the operational administration schedule and is prepared to collaborate with PARCC and its other vendors to appropriately manage the workflow to align with the goals for the operational administration.

Early in the process we provide a print materials plan and the Print Specifications by posting to the PARCC SharePoint site for approval by the PARCC. These plans will specify the type size, style, ink and paper color, paper quality, and final layout of operational administration materials. PARCC and its representatives will then review and provide input and approvals and we will make changes as needed.

Our plan will provide for collaboration and feedback at every stage of the print production process. We will use PearsonAccess to collect enrollment and student registration information. We will provide data file layouts for district and schools to upload their organizational and student data. We will work closely with states to obtain a list of current names, addresses, email addresses, and phone and fax numbers of the school and/or district assessment coordinators and information for school staff responsible for testing. Alternatively, states may upload school and district information for the entire state, instead of individual schools or district uploads.

This information will be used to setup the registration system for districts and schools to enter their enrollment information for both print and computer-based administration. We will closely monitor and review the information for completeness, accuracy, and timely submission the material's ordering window. Final print quantities and enrollment information will be provided by districts and schools and this information will provide final production quantities for printing materials, shipping materials to districts and schools, and generating packing lists.

The attached Materials List (4/6/14) documents our assumptions for number of Braille and large print materials required. Braille and large print test materials are included in the per student and per test (high school) pricing. Our costs also assume providing five percent overage to the districts to allow for printing errors, shipment and site shortages.

Producing Test Booklets

Pearson will carefully design and produce non-scannable materials such as test booklets to meet the needs of PARCC. We will involve PARCC staff in reviewing and approving the test booklets we produce for the Operational Administration. We will follow a proven procedure for preparing print-ready live test booklets, manuals, and other printed materials.

At each step, both PARCC and Pearson will review test booklets to verify accuracy and ease of use. To verify accuracy and consistency for test booklets we will employ a three-way check process. This includes a comprehensive checklist to check each document type for accuracy. Our program team will provide sign-off on the checklist, noting the accuracy of the documents. We have provided a detailed list of all print materials requested in this RFP in the Materials List (4/6/14),.

As part of our quality control process, we will send test materials through three different groups for review before sending them to the PARCC, as follows:

- Proofreading resources in our Digital Composition and Publishing group
- Content experts familiar with PARCC content
- Program team staff fully dedicated to PARCC

Each of the above-mentioned groups will perform checks to verify final forms meet PARCC's requirements. As an additional quality control step for test forms, Pearson requires a content expert not familiar with the items to take the test for paper and online forms. For the paper forms, we will send a printed sample to the key review content group. For online forms, we will publish the forms in the quality control environment and provide test tickets to the key review content group so they can take the tests. These content experts will take the tests with no knowledge of the items or answers and execute a comparison against the answer keys.

The Key Review staff will note any discrepancies on the tracking form and communicate them to the PARCC program and content teams, who then will discuss the findings with the PARCC. If corrections to a final version prove necessary, we will send an updated version for PARCC review and approval before sending to the printer or publishing to production. We recognize the need to work with PARCC staff on making changes to items at all stages. The risks of such changes will be clearly articulated so the PARCC can make informed decisions.

The following figure lists the proofing steps each PARCC Operational Administration test booklet will go through as part of our forms design quality control process. Our process incorporates concurrent procedures for verifying the alignment of test booklets, answer documents and administrator manuals—all essential to a smooth test administration. This is the standard process used by Pearson in proofing test booklets. We typically execute this process with a state department of education (note "Department" in the figure). In this case the approvals would come from PARCC.

Test Booklet Proofing Steps

Step 1
We will review the items/graphics for completeness and verify the unique number for each item.

Step 2
We will send gallery proofs of formatted items—one per page—to the Department for final review and approval.

Step 3
With the approval of the Department, we will create a mock-up of the document from the item matrices.

Step 4
We will build bookmaps for each respective test booklet and review page layouts (i.e., spacing, item order, option designation, page numbers and section are).

Step 5
One of our graphic designers will create the document and review it using a detailed checklist.

Step 6
Pearson proofreaders will follow a detailed checklist to review the document, use matrices for comparison, and mark any revisions.

Step 7
The Department and Pearson will have two cycles of proofs to review, revising the paper proofs until the document is determined to be clean and final.

Step 8
After the final review, we will create a PDF of the final proof.

Step 9
We will check the directions in the answer documents and administration manuals for format, grammatical correctness, and proper use of convections.

Step 10
Our design team will incorporate revisions to items and graphics, as necessary.

Step 11
After the review criteria have been satisfied, we will give the printing plant approval to print.

Test Booklet Proofing Steps. We will follow the detailed process above to produce accurate test materials for the PARCC Operational Administrations.

A final digital proof will be released and declared “clean” only after all review criteria have been satisfied. Only then will the document be promoted to print production. In the event there are defective materials, Pearson will replace these at least 10 days prior to the administration window.

Fully Consumable Test Booklets versus Scannable Answer Documents

Fully consumable test books require unnecessary scanning of sheets that do not contain student responses, which drive up costs and require more time in the schedule. Our recommendation is for PARCC to use separate scannable answer documents and test booklets for grade 4 and above to reduce scanning hours during data collection. Additional savings are realized during the handling and preparation of documents for scanning, and faster slitting rates occur when there are fewer sheets within a document. Scannable documents are overall more expensive to print and because test booklets have substantially more pages than answer documents this drives up the print costs. It is more cost effective to use separate scannable answer documents in terms of both printing and processing.

For each product produced, we will provide the Partnership Manager with 3 digital proof copies. Following PARCC's final approval the document will be sent to print production. During printing, Pearson will deliver 10 advance copies of each product from five equally spaced points in the print line to PARCC.

We will produce one form of each document in large print in a minimum of 18-point font on 14 x 17 approved paper and in braille. We have provided the detail around all printed materials, including regular books, braille, and large print versions along with braille notes for the braille versions in the Materials List in the **Other Supporting Material** section. This also shows were we will print a 5 percent overage of all books to allow for any printing errors, and shortages in districts and schools.

Pearson will also print any materials necessary for the operational assessment such as transmittal, memoranda, pre-ID labels, packing labels, packing lists, bill of lading and return labels. We will also supply PARCC with the layout of the packing lists and incorporate any changes.

Pearson will provide the following for all print test booklets, these specifications are also shown in the Materials List (revised 4/6/14):

- Scannable booklet covers for consumable grade 3 booklets will include demographic grids and other data fields to meet PARCC's specifications
- Scannable answer document covers for grades 4 and higher will include demographic grids and other data fields to meet PARCC's specifications
- Unique security barcodes on all secure test booklets
- 45# caviler paper or PARCC approved equivalent for cover and interior pages for non-scan print versions
- 60# white opaque cougar or PARCC approved equivalent paper for all scannable documents (we recommend 60 pound ScanTech for scannable documents) (Grade 3 test booklets and all answer documents). Answer and test documents will have matching colors
- Print covers in 1 color of ink plus black with different colors for each grade and subject to be approved by PARCC, inside test black print
- Select interior pages in one color ink plus black

- Test books and manuals size are 8.5 x 11 saddle stitched or perfect bound
- 1 PARCC approved mathematics reference sheet per mathematics test
- Randomly spiral forms
- Color-coded or unique identification marks on spines to easily identify booklets when stacked
- Shrink-wrap test booklets in standard units of 5 and 20 as determined by the number of forms
- Uncollated units for students with special read aloud accommodations in small groups
- Manipulatives (all will be provide in braille and large print as directed by PARCC):
- Ruler—1/2 and 1/4 , .5 or .1 centimeter marking, Grade 3, one per student
- Ruler—1/8 and .01 centimeter markings minimum, Grades 4-5, one per student
- Ruler —1/16 and 0 .1 centimeter markings, Grade 6+, one per student
- Protractor—Clear velum or rigid clear plastic, Grades 4-5, one per student

PARCC will have input into the design and layout of the packing and shipping lists.

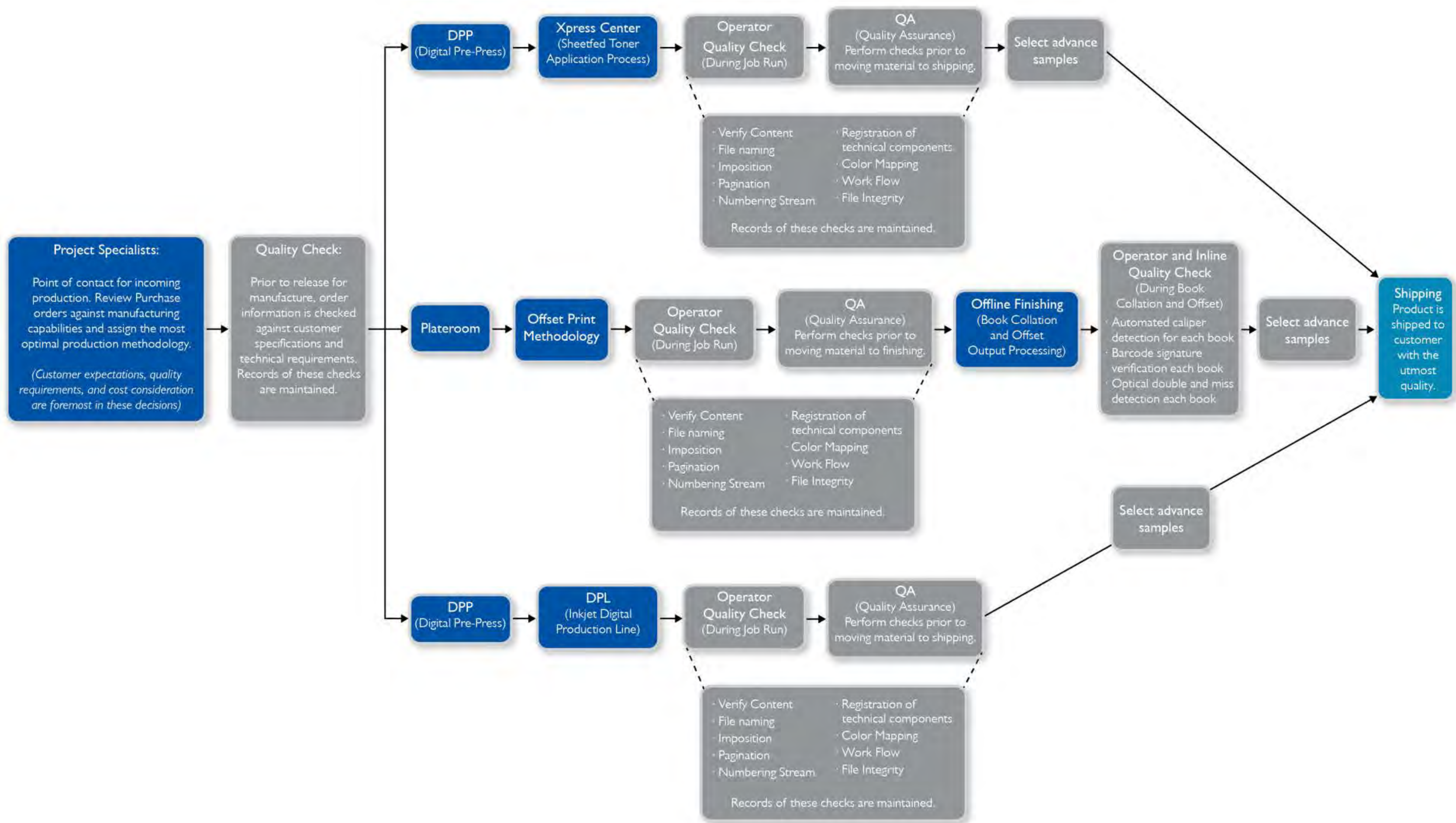
Print Quality Program

Pearson Print Service operates within the sanctions of an ISO 9001:2008 Quality Management System, and practices process improvement through Lean principles and employee involvement.

Raw materials (paper and ink) used for scannable forms production are manufactured exclusively for Pearson Print Service using specifications created by Pearson Print Service. Samples of ink and paper are tested by Pearson prior to use in production.

Project Specialists are the point of contact for incoming production. Purchase orders and other order information are reviewed against manufacturing capability and assigned to the most optimal production methodology. PARCC expectations, quality requirements, and cost considerations will be foremost in these decisions. Prior to release for manufacture, order information will be checked against PARCC specifications, technical requirements, and other communication that maintains expected outcomes. Records of these checks will be maintained.

Files for image creation will flow through one of two file prep functions: Digital pre-press (DDP) for digital print methodology, or plateroom for offset print methodology, as shown in the figure below:



Project Specialists:
Point of contact for incoming production. Review Purchase orders against manufacturing capabilities and assign the most optimal production methodology.
(Customer expectations, quality requirements, and cost consideration are foremost in these decisions)
Quality Check:
Prior to release for manufacture, order information is checked against customer specifications and technical requirements. Records of these checks are maintained.
DPP (Digital Pre-Press)
Xpress Center (Sheetfed Toner Application Process)
Operator Quality Check (During Job Run)
QA (Quality Assurance) Perform checks prior to moving material to shipping.
Select advance samples
Verify Content
File naming
Imposition
Pagination
Numbering Stream
Registration of technical components
Color Mapping
Work Flow
File Integrity
Records of these checks are maintained.
Plateroom
Offset Print Methodology
Operator Quality Check (During Job Run)
QA (Quality Assurance) Perform checks prior to moving material to finishing.
Offline Finishing (Book Collation and Offset Output Processing)
Operator and Inline Quality Check (During Book Collation and Offset)
Automated caliper detection for each book
Barcode signature verification each book
Optical double and miss detection each book
Select advance samples
Shipping Product is shipped to customer with the utmost quality.
Verify Content
File naming
Imposition
Pagination
Numbering Stream
Registration of technical components
Color Mapping
Work Flow
File Integrity
Records of these checks are maintained.
Select advance samples
DPP (Digital Pre-Press)
DPL (Inkjet Digital Production Line)
Operator Quality Check (During Job Run)
QA (Quality Assurance) Perform checks prior to moving material to shipping.
Verify Content
File naming
Imposition
Pagination
Numbering Stream
Registration of technical components
Color Mapping
Work Flow
File Integrity
Records of these checks are maintained.

Both DPP and plateroom verify content, file naming, imposition, pagination, numbering stream, registration of technical components, color mapping, workflow, and file integrity. Records of these checks will be maintained.

Offset production requires printing that uses a lithographic process. Offline finishing activities are required to create books and package offset output. Digital output may flow through an inkjet Digital Production Line (DPL) or a sheet-fed toner application process in the Xpress Center. Digital output is capable of inline finishing.

A battery of quality checks will be performed by employees in these areas. The checks will include color match, correct file selection, content match to proof, litho-code to serial number synchronization, registration of technical components, ink density controlled by densitometry, inspection for print flaws, perforations, punching, pagination, scanning requirements, and any unique features specified for the order. Records of these checks and samples pulled from planned production points will be maintained.

Offline finishing will include cutting, shrink wrapping, folding, and collating. The collation process has three robust inline detection systems that will inspect each book for:

- Caliper validation that detects too few or too many pages. This detector will stop the collator if an incorrect caliper is registered.
- An optical reader that will only accept one sheet. Two or zero sheets will result in a collator stoppage.
- The correct bar code for the signature being assembled. An incorrect or upside down signature will be rejected by the bar code scanner and will result in a collator stoppage.

Our Quality Assurance (QA) department personnel will inspect print output prior to collation and shipment. QA also supports process improvement, work area documentation, audits process adherence, and establishes training programs for employees.

Requirement

V.B.2.G. Pre-Identification of Consumable Test Booklets and Answer Documents

Response Requirements for Section V.B.2.G.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.G

Deliverables for Section V.B.2.G.

a) Deliver pre-Identification Specifications

b) Deliver pre-Identification Labels

Pre-Identification of Consumable Test Booklets and Answer Documents

Pre-ID Labels

Pre-coding or pre-Identification (pre-ID) of labels with student-level data is an excellent way to save school personnel and students' valuable time that would otherwise be spent gridding demographic information onto scannable documents. Pre-ID labels also streamline the entire distribution and collection process, improving data collection accuracy. To save time for test administration, we will provide labels that are pre-identified with machine and eye-readable student names and other identifying information.

For any new students or students that were otherwise not included in the demographic file upload or manually entered into the system before the labels were created they can use a generic pre-ID label or hand-grid the demographics page, For pre-ID label's, program information is pre-printed; however, the information can be updated by hand-gridding at the test site to correct or by updating the student information within the student registration system. Pearson will package pre-ID labels and deliver them to each designated site as part of the shipment containing test administration materials. We will also provide pre-assigned login IDs and passwords for test day administration for online testers based on the information gathered for pre-ID.

Specifications

Pearson will provide the pre-Identification specifications to PARCC six months prior to each Operational Test administration window for review and approval. These specifications will include proposed milestones for specific pre-Identification activities (e.g., data update deadlines, label production, delivery to districts and schools).

Pearson will also include samples of all memoranda required in the pre-identification process for PARCC approval. We will provide a quality control plan to PARCC two months prior to the operational administration. The plan will confirm that the labels delivered to districts contain valid district data, the options selected by districts, and a level of print quality to enable accurate scanning and preclude the possibility smudging, smearing, and/or flaking.

Pearson will confirm receipt of student files from PARCC states and that upon submittal, the pre-ID file passes through an edit check before it is loaded into PearsonAccess. Another assurance check is the resulting confirmation message provided to the submitter indicating whether the file processed successfully or if errors were encountered. Records with suspect data will be indicated in the messaging to the submitter so that appropriate updates can be made and verified in collaboration with school staff.

Pre-ID System

PARCC will have a single data source for uploading and managing student pre-ID information through the administrative system. PearsonAccess will be used all four years of the contract for districts and schools to upload and manage student demographic information.

The administrative system, PearsonAccess, will collect student data based on the agreed upon requirements via a data file upload, or manual entry into the system. Then, authorized users can add, modify, and manage student demographic information that has been successfully uploaded to the administrative system. This includes demographic information associated with students who are excused or exempt from testing. PARCC districts and schools will have access to review and manage their own data, which supports data accuracy and shortens response time. We will include system alerts so users are aware of possible key entry or transcription errors when they occur—saving end users time and improving accuracy.

Pearson will work PARCC to determine the best file formats to accommodate the state and school districts, typically.csv and .txt files. Student data will conform to the PARCC data interoperability formats and interoperability guidelines. We will work with PARCC design the student pre-ID file layout and obtain PARCC's approval on the final layout and pre-ID plan. This plan will be designed to help minimize confusion and enable end users to efficiently sort students from pre-load files into testing sessions.

We will provide pre-ID labels in conjunction with a pre-ID roster, organized by building or by classroom, depending on school preferences. The same stringent quality processes apply to pre-ID printing, whether printed directly onto answer documents or onto labels. We will also provide a separate roster of students using the same sort order as in the pre-ID labels so districts and school can verify the accuracy of their information before the label is used.

Training

Pearson will train districts and schools on the procedures for submitting and validating their data prior to actually opening the window for data uploads via the administrative system. As part of the "Administration of Paper-based Assessment" and "Administration of Online Assessments" training modules described on page V.B.43-we will provide an online tutorial for districts and schools that provides a detailed step-by-step process to collect and upload their pre-ID information. In addition to training we will provide an online checklist for districts and schools to use in preparing their pre-ID file submission. This will confirm users are fully trained and familiar with the process prior to using the system. Pearson will closely monitor the progress of data submission and run systematic quality checks to confirm all the data fields were properly submitted and are complete. We will confirm the data submission to districts and/or schools including information that describes the number and types of students by school submitted the order in which we will print the labels, and the range and incidence of values in selected fields.

Quality Control Plan

Pearson will develop a quality control plan to cover proper training documentation for using PearsonAccess The plan will cover the entire pre ID process and include file creation, submission, validation of data, label printing, and application. We will provide the plan at two months prior to the operational administration.

Requirement

V.B.2.H. Mock Data End-to-End (Test Deck)

Response Requirements for Section V.B.2.H.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.H

b) Offerors will propose comprehensive solutions for conducting Test Deck reviews.

c) Offerors may propose an alternate mock data review process if significant schedule and/or cost benefits would result from a different approach.

d) Offerors must include budgeted amounts in their Price Response/Reply and Budget Worksheets to cover the all reimbursable costs for all participants' air travel, car rental, lodging, and miscellaneous expenses such as parking, tolls, vicinity mileage, and other costs Offerors identify as necessary to participants' travel. Participants will also be reimbursed for meals at the GSA per diem rate.

e) As an "add on", the Offeror shall provide a price for individual states to include sample documents and/or online responses for their own review: The cost estimate shall include the following for one state:

a. Total mock student count will not exceed 210 students

i. 5 computer-based students per grade per subject/course

ii. 5 paper-based students per grade per subject/course

b. The district count will not exceed two districts.

c. The school count will not exceed two schools per district.

d. The state shall receive all materials necessary to code student answer documents (e.g. keys, pre-identification labels, headers)

e. The state shall submit student and organization registration data

f. The state will send paper documents for scanning and scoring

g. The state shall input student responses for computer-based test cases

h. The Contractor will ship scanned documents and processing reports back to the state

i. The Contractor will provide a student data file to the state.

Deliverables for Section V.B.2.H.

a) Deliver Mock Data Test Plan
b) Deliver Mock Data Files
c) Deliver Mock Data Results
d) Deliver TOR Meeting Plans
e) Deliver TOR Review Meeting Agendas, Participant Lists, Meeting Notes
f) Deliver TOR Review Training Materials

Response

Mock Data Test Plan

Two types of testing (End to End and Customer Acceptance Testing) use mock data to validate that Pearson products are operating according to agreed-upon specifications and to verify that the requirements are valid and reflect customer expectations. Pearson agrees to the requirements for mock data testing as stated in the RFP and will deliver a mock data test plan, mock data files, and mock data results to PARCC for review.

Additional details about this process are described in this section, and the mock data test plan is based on work completed at a consortia level. We acknowledge the request to allow states to include sample documents and/or online responses; however, the pricing sheets did not provide a means to include these add-on costs. With additional instructions we can provide separate add-on costs during the time of oral presentations or negotiations, if requested.

The following assumptions were used for the Mock Data End-to-End testing:

- Testing would be complete at the consortium level for each administration, not each individual state. One state would be chosen to run using the test cases.
- We expect to have requirements for end-to-end testing six months prior to test start date.
- End-to-end testing will be performed in a test environment and we expect that a complete environment will be available to complete the testing.
- Security and performance testing are not part of the end-to-end testing cycle. It is expected that the yet to be determined, Technical Operations vendor would be responsible for security and performance testing.

End-to-End (E2E) or Acceptance Test. Pearson testers will replicate a set of customer-like test cases and data to introduce and process through the systems similar to how PARCC's data is expected to traverse. This testing is intended to replicate PARCC's experience using the same applications and user interfaces that the customer would use, but with a much smaller set of data than would be expected in production. Testers will use a standard set of test cases that address common packaging and distribution, processing, scoring, and reporting rules as well as test cases that address PARCC specific requirements. Testers will create organizational structure and corresponding student data as it would be created and introduced in production for both paper and online test processing. Introduced at different stages of the assessment lifecycle, the data would be expected to traverse the systems that consume/pass/merge the data throughout the downstream processing systems.

The tester will validate the PARCC-like organizational and student dataset in the test environment, at specific validation points, as it is processed through different stages (packaging and distribution, scanning, online testing, scoring, and reporting).The tester will then verify that the data is processed correctly and that results are as expected for the PARCC deliverable or at each validation point. This data is expected to stay intact and be processed throughout the entire PARCC path, just as the data would be processed for PARCC in production. This type of testing is considered to be black box testing, where inputs to and outputs from specific systems or applications are used as checkpoints for data validity along an entire process.

The end-to-end check will include:

- A minimum of 150 mock computer-based test cases per grade level per subject for each assessment for verification of the following:
    - Capture of correct machine-scorable (MS) information for every form
    - Confirmation of capture of the correct constructed response (CR) area
    - Correct importing of student demographic information
    - Accurate merging of MS and CR scores
    - Scoring keys and scoring programs are correct
    - Test disruptions are handled correctly
- A minimum of 150 mock answer document test cases per grade level per subject for each assessment for verification of the following:
    - Each grid area is being properly scanned and recorded for every form
    - Each respond area is being properly scanned and recorded for every form
    - Student demographic information, including pre-identification files are validated
    - All cases involving missing and incorrect information is checked
    - All cases requiring editing are properly related to an editor
    - Procedures to assure accuracy of data process are validated
    - Scoring keys and scoring programs are correct
    - Not tested records are dealt with appropriately

Pearson may use computer-generated data to supplement the test set. For each administration, a data file will be provided that verifies that student answer documents and computer-based scoring responses have been scored accurately. Additionally, the data file will allow the Partnership Manager and staff to monitor the end to end processing of student responses until completion.

Customer Acceptance Testing (CAT). The Pearson test engineer will incorporate PARCC defined test cases (mock data) into its end-to-end testing. This will allow PARCC to provide input into how the testing is performed and/or what data is used in the testing process. CAT is designed for the customer to contribute to and participate in testing Pearson products in Pearson test environments before the products are delivered in production.

CAT will enable PARCC to validate that the Pearson products and processes are functioning according to the agreed-upon specifications at distinct stages in the process. PARCC can also use CAT to discern whether the requirements are valid or reflect their expectations. PARCC may specify the test cases, processes, or deliverables they would like to validate. Those deliverables could be contained in a user interface, packaging and distribution, scanning, online testing, scoring, and/or reporting.

Customer involvement in test stages allows corrections to be made earlier if discrepancies are found, which can greatly contribute to the success of the program. This testing is usually performed in conjunction with Pearson's end-to-end acceptance test group.

Test Deck Review Meeting

Prior to each administration window (up to three times per year) Pearson will facilitate a Test Deck Review Meeting. During this meeting PARCC representatives will review all of the findings from the mock data end-to-end test.

Prior to each meeting, Pearson will hold our own internal pre-review of the mock data. PARCC will assemble a Test Deck Review (TDR) Committee including 10 state departments of education representatives or their designees. One committee member will serve as the lead and at least one member will be an accessibility expert. The meetings will be held at one of Pearson's facilities. For costing purposes we assumed the meetings would be held at our Iowa City/Cedar Rapids offices.

Participants' Travel Costs

As described in section **V.A Test Development**, the cost of participants' air travel and/or rental car (as needed), lodging, meal per diem, and miscellaneous expenses such as parking, tolls, mileage, and other costs are included in our Cost Proposal. Pearson will also be responsible for all meeting room and equipment costs.

At least eight weeks prior to the Test Deck Review Meeting, we will work with the Partnership Manager to identify PARCC and PARCC State Department of Education K–12 staff or their designees who can attend the meeting at Pearson. We will also provide a Test Deck Review Meeting Plan, which will include the meeting agenda, participant list (to be updated prior to the meeting with final list of DOE staff that plan to attend), and other required information. All transportation and lodging arrangements will be made through Pearson and direct billed for participant convenience. Reimbursement for per diem expenses will be provided within four to six weeks following the Test Deck Review Meeting.

Below are assumptions regarding the Test Deck Review Meetings that will be held each year. All Test Deck Review Committee meetings will be in-person. This information was provided in the "Travel and Meetings" tab of the Cost Proposal template.

In-Person Meetings Per Year	Meeting Duration Per Meeting (Days)	# of State Participants Per Meeting	# of PARCC Staff Attending Per Meeting	Total Travelers Per Year (State + PARCC)	Number Participants Eligible for Stipends/Substitutes per Year	Location
3	5	8	2	30	0	Contractor Site

Requirement

V.B.2.I. Packing, Distribution, and Retrieval of Test Material

Response Requirements for Section V.B.2.I.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.1

Deliverables for Section V.B.2.I.

a) Deliver Packaging, Distribution, and Retrieval Specifications
b) Deliver Missing Secure Materials Report
c) Deliver Braille Materials
d) Deliver Large Print Materials
e) Deliver Test Materials

Response

Pearson acknowledges the importance of packaging high-stakes testing materials accurately with on-time delivery to districts or schools. Pearson has three facilities—Cedar Rapids, Iowa, Austin, Texas, and Owatonna, Minnesota—to support PARCC packaging, distribution, and retrieval specifications. All our packaging facilities are certified to ISO 9001:2008 Quality Standards and processes and provide consistent, repeatable, and proven packaging processes.

The packaging specifications will be delivered two months prior to test administrations. The information includes detailed procedures for how we will package and distribute materials to districts and schools, examples of packing lists, and the information and materials needed for return of materials back to Pearson for inventory, processing, and storage. The plan will include samples of memoranda required for the delivery and pickup process.

The plan will also include information about the following:

- Size and weight of shipping boxes
- Specifications for shrink-wrap materials
- Labeling of shipping boxes
- Barcoding of materials
- Use of any additional materials such as a map listing the identity and location of boxes on each pallet
- Proof of delivery
- Materials overage

Pearson will obtain PARCC approval on the specification plan, district/school communications, and memoranda prior to the start of the packaging process. The Partnership Manager will receive examples of all materials, including any materials associated with computer-based testing. When ready to ship materials, Pearson will seek Partnership approval before allowing any materials to be sent to schools or districts. Additionally, Pearson will seek Partnership approval prior to providing computer-based testing access.

Recognizing that DC, IL, MA, MD, and RI require their materials to ship directly to schools we will meet the unique needs of each state by setting up packaging and distribution specifications at the district or school level to accommodate a state's preference.. Pearson will pay for all charges associated with shipping to and from the districts and schools. In the event there is a delivery error, Pearson will quickly take action to rectify the situation, and overnight replacement materials.

To confirm the accuracy of our packaging of each order, we offer automated quality control verification that accounts for materials in real-time as the unique barcode on each packaging component is picked and scanned. Our solution provides PARCC states with detailed confirmation that we have packaged each order accurately.

Quality control occurs throughout our material packaging, distribution, and retrieval processes. From collecting enrollment counts to distributing and receiving test books and answer documents, we will document and follow quality control procedures and checklists.

Packaging Specifications

The following processes will be followed to make sure that test materials are distributed properly:

- During unit and product configuration testing, the software developer and IT assessment testing group will perform extensive unit and product configuration testing on distribution software programs. This includes verifying that our solution accurately collects enrollment counts and student information, and that we have properly calculated the material distribution rules.

- Before each administration, our Organizational Quality group will perform acceptance testing, which includes entering enrollments and submitting student information as a mock district and processing that enrollment information through the packaging of materials in the exact manner as for live data. This activity verifies output, confirms integration of products, and verifies that distribution materials meet customer expectations.

Pearson will use the following production validation steps to package and ship live materials correctly:

- A pre-blue dot checklist will verify completion of pre-production activities before blue dot production and confirms the operations departments' preparation for printing and packaging the test materials.
- The pre-blue dot production process will use a small sample of carefully selected districts to check for key packaging, processing, and reporting characteristics. During this process, we will verify distribution materials for accuracy, completeness, print quality, and adherence to PARCC requirements.
- Random spot-checking will occur during packaging to verify that we adhere to specifications throughout distribution.

The following figure shows our quality assurance process for materials distribution.

Materials Distribution

1. Unit and Production Configuration Testing

- Performed on all distribution software.
- The testing includes verifying that enrollment counts and collected student information are accurate.
- Verify all distribution rules are calculated properly

2. Acceptance Testing

- Performed by independent Quality Group

3. Product Validation

Pre-Blue Dot

- All pre-production activities are complete
- All operations departments prepared for printing, packaging production materials

Blue Dot

- Specific districts containing pre-determined criteria are processed
- Distribution materials are verified for accuracy, completeness, print quality
- Verify distribution materials meet customer expectations

Packaging Spot Check

- Random sampling during packaging to verify adherence to specifications

Verifying Materials Distribution Accuracy. We use a series of quality control steps to verify the accuracy of our materials distribution.

Our plan for the distribution and return of materials focuses on maintaining test security, providing accurate handling of all test assessment materials, and delivering these materials to the participating schools in a timely manner.

Our experienced warehousing and transportation departments will maintain the quality and security of material distribution and return by using such methods as sealed trailers, and hiring a reputable carrier with the capability to immediately trace shipments.

To maintain the greatest degree of test security, accurate handling, and timely delivery, Pearson will:

- Create a transportation file consisting of requested quantities of each material type (secure or non-secure) along with calculated overages for each participating school within each district
- Use internationally-accepted best practices through Pearson's Oracle packaging and distribution application
- Identify appropriate shipping modes based on shipment size and destination
- Verify shipping addresses for validity according to carrier files
- Package those materials intended for a site in boxes addressed to that site
- Ship materials directly to the district coordinators of the PARCC assessment who will in turn be responsible for distribution of all testing materials
- Construct a packaging and distribution schedule to provide that districts receive testing material at least two weeks prior to testing
- Use our distribution system's tracking capabilities to provide precise status information and immediate opportunities for corrective action

We will make this tracking capability service available in the following ways:

- Working with UPS, or other approved vendors who provide receipted delivery and traceable motor freight
- Using barcode labels on every package to make shipping and tracking packages faster and easier
- Managing complete shipment accountability from origin to destination, using our in-house tracking system as well as online connections with the carrier
- Generating pallet maps that help locate boxes on that pallet for ease in distributing materials to schools
- Generating district packing lists showing exactly which materials are in each box and how many boxes should be received
- Providing school packing lists and return shipping labels and forms to each test coordinator

Distributing Materials to PARCC States

Pearson transportation specialists are adept at planning major distributions of assessment materials for numerous state departments of education, school districts in major urban area settings, and remote locations within states as well as international destinations. As a result, our major carrier, UPS, is experienced in distributing critical assessment materials and knows the importance of securely transporting the materials so that no assessment content is compromised.

As soon after contract award when state participation and enrollments are communicated to Pearson, distribution points will be documented and researched as part of the packaging and distribution planning. Estimates for shipment size will be developed, and help identify the mode of transportation that is appropriate to circumstances at each destination. Pearson typically uses a number of shipping modes, including parcel for small freight, and less than truckload (LTL), or full truckload carriers for larger states, schools, and districts. Parcel carriers are best suited for units less than 150 pounds and are frequently used for additional orders, or back ordered and replacement items.

Districts that do not have adequate warehouse or dock facilities may require assistance with materials handling at the time of delivery. Those districts will be identified during the planning process and special accommodations will be made by the delivery team.

Prior to finalizing distribution plans, we will present all optional carriers and solutions for approval by PARCC. Pearson will implement a plan to correct any shipping or order errors and to locate and replace any missing shipments. As part of our planning documentation we will also agree to:

- Alert the Partnership Manager of all shipping related issues
- Trace any missing shipments to resolution using all available tracking systems
- Ship additional or replacement materials via the most appropriate and timely method (second day or overnight delivery as needed) so the test administration date is not compromised
- Maintain an error log of all issues and deliver in both paper and electronic format to the Partnership Manager three weeks after the scheduled delivery

Efficient Packaging Processes

Pearson’s extensive and continuing investment in improving our packaging and distribution processes has resulted in a system that drives accuracy, cost efficiency, and smooth test administrations.

Two primary features of our packing and distribution system deliver accurate packing for even the largest and most complex assessments:

1. Unique barcodes identify each item, box, order, pallet, and shipment.
2. Packing lists are printed after the order is filled, based on barcode scans made as materials were packed.

Barcodes for Material Tracking

Our barcode system is an essential component of Pearson’s established packaging and distribution process, for a secure, accurate, and efficient test administration.

We use barcodes to identify materials, track them within our system, and for accurate packaging, labeling, and shipping.

We can locate PARCC materials at any step in processing, shipping, return, scoring, and storage. With these quality control steps, we achieve extremely accurate results for very large and complex assessments.

Below is a description of how this system will be used for the PARCC Assessments:

- Before packaging, we will print a unique barcode serial number on secure test booklets and other PARCC materials that require security.
- We will use an inline quality control system to verify that barcodes are printed sequentially on like materials before they are packaged (shrink-wrapped) in manageable packet sizes (for example, fives or tens).
- Barcode ranges will be recorded on ID sheets to be included in the shrink-wrapped package of test booklets. An electronic copy of the barcode ranges will be available to for the district/school to view in PearsonAccess.
- District materials will have contiguous barcodes within form types. This provides easy and correct material check-in and verifies the return of materials sent.
- Customized items such as pre-Identified headers, shipping labels, return materials, and pre-identification labels will have barcodes that identify the type of item and the specific order to which each item belongs so the right materials go to the right destination.
- We will store PARCC barcode data in a computer master file and print it on packing lists and security reports.

3250 115 -9

Barcode Security. Unique barcodes identify each item, box, pre-packed component, pallet, and shipment. We can track your secure materials at any stage of the process.

Pick and Pack Process for Accurate Packing and Documentation

Pearson's order fulfillment is guided by system-generated pick lists detailing the numbers and types of materials to be shipped to districts. Our attention to detail gives PARCC states' confidence that each school or district will have the right materials delivered on time to ease test administration.

The following steps explain the pick and pack process that will be used for PARCC Assessments:

- We will assign bar code ranges by form type and give each district contiguous barcode ranges to expedite check-in of test materials.
- We will use handheld radio frequency (RF) scanners with a system-generated pick slip to select the materials required for each state's school or district.
- Operators will scan the barcode on each item as it is placed in the box to fill an order.
- If an operator scans an incorrect item or quantity, the system will generate a report and a scanner lockout. A supervisor must correct any inaccuracy before the operator can continue.
- The system will perform a quality control check to confirm we have packed items to the pallet in the correct order or that we have them on back order.
- Once materials have been packed, our quality control system will generate a packing slip and pallet detail report for each shipment. Quantities on the packing list must match exactly what was prepared for shipment before order can be shipped. This feature greatly facilitates the identification, management, and distribution of test materials once they have been received by districts or schools.
- If an error occurs during the final packaging process, the system will generate an error report. Documents cannot be shipped to print until the issue is resolved.
- Pearson personnel will track shipping and order status online.

Our automated quality control verification accounts for materials in real-time as the unique barcode on each packaging component is picked and scanned.

Documentation generated from the system or provided in handbooks prior to test administration will tell test coordinators and administrators exactly what they need to do when they receive the test materials, administer the tests, and return the completed documents.

Identifying and Packaging Accommodated Test Materials

Our processes for accurately packaging secure test materials such as braille, large-print, and audio or video formats include applying a secure barcode for accountability purposes. We will print the number on a label and apply it to the item prior to final packaging. We will assemble these materials in packages of various types and sizes and they will be matched up with the boxes of regular test materials, so they will all arrive in the same shipment.

Electronic Tracking. Pearson's packaging software has checks in place to prevent an operator from packing the wrong material.

Packaging Non-Secure Items Unique to Your Administration

Pearson processes save you time, but not at the cost of accuracy. Just prior to pre-assembly and packaging, non-secure materials that do not have a secure barcode on each item will be printed with a unique barcode on the assembly identification sheet.

For instance, non-secure items—such as formula sheets or graph paper—will be assembled in various, specific package types and sizes to prepare them for final packaging. The barcode on the assembly identification sheet will identify all the items within that assembly making it easier for coordinators and administrators to account for materials. The number will also validate that all items are included in the boxed materials.

To standardize assessments across schools/districts, Pearson will also provide protractors and rulers, as shown on the Materials List (revised 4/6/14).

Additional Shipments

Last minute requests for additional materials necessitate quick response times without compromising the security of the test materials. When a district or school identifies any shortages, the test coordinator or administrator can request materials in one of three ways:

1. Internet—enter shortages into the Administration Portal
2. Email via a link provided in the Administrative Portal or by using the PARCC Operational Assessment Materials email address
3. Additional Materials Request Form—fax a form to Pearson's project team
4. Toll-Free Call—contact your Pearson project team or Customer Service Center

Orders received prior to 11 AM CST are usually shipped the same day; all others no later than next day. Overnight delivery is planned for most additional orders, with ground shipment making up the majority of the original orders.

Retrieving PARCC Assessment Materials

Pearson will be pre-pay all charges associated with returning materials from the districts and schools. The necessary return materials will be provided in the Materials Return Kit:

- Return labels
- Prepaid postage labels
- Freight bills-of-lading
- Instructions

We will include step-by-step instructions so districts and schools can easily assemble, box, and label used testing materials for collection after test administration. We will provide double-column boxes with extended flaps for distribution and return of materials. These boxes have a maximum bursting strength of 275 pounds.

Pre-printed mailing labels will be provided for the return of materials after the administration. These color-coded labels (one unique color for scorable materials and another unique color for non-scorable materials) are used by Pearson receiving staff to confirm that the number of boxes the district indicates they shipped matches the number of boxes we receive and quickly differentiate between scorable and non-scorable materials, which are packaged and returned separately. Paper bands for bundling and returning test documents are also provided as part of the materials return kit.

Test coordinators will have ample materials for organizing student response documents and test booklets prior to the return of materials. Pre-paid shipping manifests, preprinted return labels, sufficient number of boxes, and easy to use instructions will make materials return less hectic following testing. The RFP states that “the Contractor will allow at a minimum, excluding district holidays, 14 calendar days after completion of the test window for district or school materials pickup.” PARCC has more recently agreed to have districts or schools return materials within 7 days after the completion of testing.

We have assumed ground return for PBA test materials and ground return for 70 percent of the EOY test materials. Up to 30 percent of the EOY materials will be overnight returned to Pearson to meet scoring and reporting deadlines.

Once returned boxes arrive at Pearson for processing, the box label will be scanned for accountability and will become part of the daily receipt log. Boxes of materials will be staged for opening by data preparation staff. As each box is opened, it will be examined for materials that are not answer documents or secure test books. Any discrepant materials will be staged for resolution by the Pearson PARCC Program Team.

If an answer document is discovered in secure materials, it will be placed in the data collection process stream for data capture. In 2012, more than 20,000 answer documents were returned with secure materials. These were quickly identified, and were processed and included in the aggregated results database for the customer.

Similarly, the program team will be made aware of any receipts that appear to be less than expected. We understand the need for 100 percent accounting for all secure materials distributed to and returned by the schools and districts.

At the completion of the administration, the Program Team will prepare a missing material report for all secure materials based on the returned materials scan file. This preliminary report will be shared with the Partnership Manager for approval prior to sharing with the districts. The Partnership Manager will have 15 working days to review. Upon approval, Pearson will notify the district and school coordinators and provide them with district and school level summary reports listing the specific missing materials, and we will work with them to locate missing materials.

All located materials will be noted and a final missing materials report will be delivered to the Partnership Manager, each state, and district. Pearson will maintain an inventory of the complete list of missing materials. The dates for these reports will be negotiated upon contract award.

Requirement

V.B.2.J. Scanning and Editing Student Responses

Response Requirements for Section V.B.2.J.

a) Offeror’s proposal shall include a response to the requirements specified in Section V.B.2.J

Deliverables for Section V.B.2.J.

a) Deliver Scanned Student Responses

Response

Assessments are all about results for the student, teacher, and stakeholders once the actual test administration is over and for the paper-based testing, this begins with scanning. Pearson will scan all scannable test books and answer documents received for each administration. The braille and large print documents will be transcribed onto regular answer documents prior to receipt at Pearson.

Accurate scoring and reporting for PARCC begins with capturing student response marks on response documents in the most efficient and timely manner resulting in reliable student data and report results. Processes and steps that contribute to our accuracy include:

- Commitment to accurate data capture and scanner improvements
- Early and rigorous testing of systems and continuous quality checks
- Application of controls and technologies for successful results
- Maintaining data integrity while data is captured and discrepancies resolved
- Storing documents in secure and traceable manner
-

Scanner Throughput Improvements. Our commitment to improve scanning rates and scanner improvements for our customers has recently improved throughput rates by 20 percent and provides consistent performance in collecting data. Pearson scanner engineers have been able to improve scanner efficiencies through diligent research, adjustments, and upgrades in anticipation of the volume of PARCC scannable documents.

Verifying Accuracy of Data Collection and Processing. Quality control processes play critical roles throughout data collection and processing. Pearson will use expected results testing methodology to verify proper collection and editing of data and to confirm that we have interpreted PARCC requirements accurately.

The following processes will take place during paper data collection and editing:

- During unit testing, we will review and approve unit test plans before executing tests, to confirm that software components are complete before we begin product configuration testing.
- During product configuration testing, we will use multiple test decks (both manual and automated) to verify proper collection and editing of data for answer documents. We will hand-grid unique scannable documents with approved documented test cases that have been standardized across administrations. Each test case will also contain an expected result. For verification, we will electronically compare the actual result against the expected result.

In addition, we create a separate test deck, using approved documented test cases standardized across PARCC administrations. We will do this to test the established editing rules. When we introduce data to the editing system, we will compare output to expected results in an automated fashion to verify proper collection and editing of data. This process also improves accuracy by reducing issues, and reduces time required for this verification.

- Acceptance testing consists of emulating the receipt of live materials to verify that we process them properly. Acceptance testing material will be processed by the operations area that will be processing the live material during production. The Organizational Quality group will verify the output to confirm proper collection and editing of data. This process also verifies that test data are processed in the same manner as live material.

Verifying Correct Capture and Editing of Live Materials. Pearson will use the following production validation steps to verify correct capturing and editing of live materials:

- A pre-blue dot checklist will be used to verify completion of pre-production activities before the blue dot process, and will confirm that the operations departments are prepared for processing live materials.
- The blue dot will consist of specified districts containing pre-determined criteria needed to provide data capture and editing quality. During this activity, the Organizational Quality group will verify proper capture and editing of data from unique scannable documents, based on a pre-defined sampling of materials. We will do this to provide accurate live processing and to confirm adherence to customer requirements.

Quality Scoring Process

1. Unit and Production Configuration Testing

Unit Testing

- All unit test plans are reviewed and approved prior to test execution

Product Configuration Testing

- Multiple test decks (created both manually and automatically) are used to verify scoring keys are being referenced accurately
  - Verify all aggregation rules are being properly applied according to the requirements
  - Each unique score point is tested independently for all unique scoring combinations
  - Separate test decks for created to test all of the aggregation and matching combinations
  - Expected results are compared to the actual output of the system

2. Acceptance Testing

- Consists of emulating receipt of live materials
- Test material is processed as live material would be by operations
- Output is reviewed by independent Assessment and Information Quality Group
- Same materials used to verify the Materials Distribution and Data Collection systems are used to verify the scoring system.

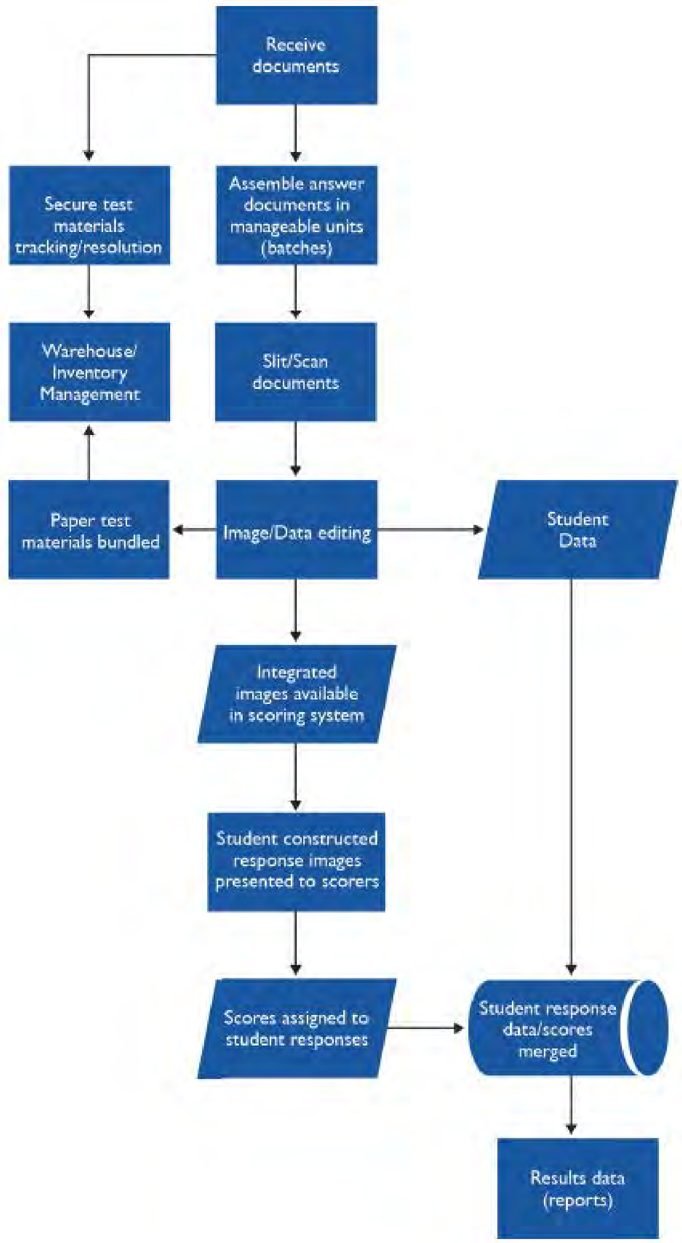
3. Production Validation

- Performed once live materials are scored
- Independent Assessment & Information Quality Group verifies all data are scored, aggregated, and matched properly.

Quality Processes for Data Collection and Processing. We use expected results testing methodology to verify collection and editing of data and adherence to PARCC requirements.

Production Scanning

Our solution for collecting data from PARCC student documents is illustrated here, and described in the text that follows.



Document data will be captured using Pearson's sophisticated "one scan" image scanning technology. This technology allows the scanning function to capture the demographic data and multiple-choice item responses while simultaneously imaging the constructed-response items to be transferred to our digital scoring platform ePEN. It establishes a direct correlation among OMR, images, and demographic data, a correlation maintained throughout the process and scoring cycle. Using these procedures contributes to meeting the requirement that all scanned responses and images are assigned to the correct students and schools.

Scanning Capabilities to Meet PARCC's Needs. Pearson maintains its own scanners with exclusive features that offer significant advantages to PARCC.

Establishing and maintaining the accuracy of our scanning, editing, and imaging processes is a cornerstone of our scoring process. While the scanners are designed to perform with great precision, Pearson exercises other quality assurance processes to confirm that the data captured from scan processing produce a complete and accurate map to the expected results. We encourage the active role many of our customers take at this point in verifying the scan file output against a representative set of test documents to confirm performance of the scan and edit programs.

Pearson pioneered optical mark reading (OMR) and image scanning, and we continue to improve our own scanners for this purpose. Software programs written by a software development team, drives the capture of student demographic data and student responses from the test materials during scan processing. Routinely scheduled maintenance and adjustments to the scanner components (e.g. camera) maintain scanner calibration. Test sheets inserted into every batch test scanner accuracy and calibration.

Controlled processes for developing and testing software specifications include a series of validation and verification procedures to confirm the captured data can be mapped accurately and completely to the expected results and that editing application rules are properly applied.

Editing Services for Accuracy Checks and Corrections

The final step in producing reliable data for scoring is the editing process. Once information from the documents is captured in the scanning process, the scan program file will be executed, comparing the data captured from the student documents to the project specifications. The result of the comparison is a report (or edit listing) of documents needing corrections or validation. Image Editing Services will perform the tasks necessary to correct and verify the student data prior to scoring.

Using the report, editors will make certain that all unscanned documents are scanned, or the data is imported into the system through some other method (flatbed scan, key entry). Documents with missing or suspect data will be pulled from the cart, verified, and corrections or additional data will be entered. Standard edits (at minimum) will include:

- Incorrect or double gridding
- Incorrect dates (including birth year)
- Mismatches between pre-ID label and gridded information
- Incomplete names

We depend on our customers to tell us what the editing rules are and we document them in project specifications for editing services. It is that set of rules and alert conditions the editor uses to determine if an error is correctable and will correct and produce clean data. If an error is determined uncorrectable, the issue will be elevated to the Pearson PARCC Program Team for resolution. When all edits have been resolved, corrections will be incorporated into the document file containing student records.

For example, if the edit listing shows a student document has the current year rather than the student's birth year gridded in the demographic area, the editor will attempt to resolve the error using information hand-printed on the student document. If the editor cannot resolve the error, the order will go on alert and the customer will be contacted for resolution.

Additional quality checks will also be performed. These include student n-count checks to make certain:

- Students are placed under the correct header
- All sheets belong to the appropriate document
- Documents were not scanned twice
- No blank documents exist

Finally, accuracy checks will be performed by checking random documents against scanned data to verify the accuracy of the scanning process.

Once all corrections have been made, the scan program will be tested a second time to verify all data is valid. When the resulting output shows that no fields are flagged as suspect, the file will be considered clean and scoring can begin. If suspects are still present, the process will be repeated before scoring begins.

Once all scanning has been completed, the right/wrong response data will be securely handed off to the PARCC data warehouse. Pearson will work with the other PARCC vendors for sharing of other related scanning data in a timely manner.

Requirement

V.B.2.K. Scoring

1. Key-Based Scoring
2. Rule-Based Scoring
3. Handscoring
4. Rangefinding and Rangefinder Review Meetings
5. Produce Handscoring Materials
6. Contractor Staffing for Handscoring
7. Team Leaders
8. Recruit and Hire Readers
9. Training and Qualifying of Readers
10. Handscoring Reports
11. Scoring Student Responses
12. Monitor and Maintain Handscoring Quality
13. Special Scoring Requirements for New York

<u>Deliverables for Section V.B.2.K</u>

a) Deliver Scoring Specifications
b) Deliver Handscoring Specifications
c) Deliver Rule-Based Frequency Distribution
d) Deliver Rangefinding Responses
e) Deliver Scoring Guides
f) Deliver Training Sets
g) Deliver Qualifying Sets
h) Deliver Validity Sets
i) Deliver Machine-Scored Student Responses
j) Deliver Handscored Student Responses
k) Final scoring plan for how scoring will be conducted in New York and implementation of the plan

Response

Scoring

PARCC forms will include a demanding and complex array of item types, as shown in the figure below, and require accompanying scoring rigor for rule-based scoring, key-based scoring, and handscoring.

PARCC Item Type	Valid Paper Type	QTI Interaction Type	ELA Item?	Math Item ?	Grade Band 3–5	Grade Band 6–8	Grade Band HS	Score Type
Category 0 Non-TEI								
Selected Response	Y	Choice Interaction	Y	Y	Y	Y	Y	Key-based scoring
Gridded Response	Y	Text Entry Interaction	N	Y	Y	Y	Y	Rules-based scoring
Multiple Select	Y	Choice Interaction	Y	Y	Y	Y	Y	Key-based scoring
Category 1 Non-TEI								
Constructed Response	Y	Extended Text Interaction	Y	Y	Y	Y	Y	Handscoring
Constructed Response w/EE	Y	Extended Text Interaction (class = tei-ee)	N	Y	Y	Y	Y	Handscoring
Fill in the Blank—Specific Character Set	N	Text Entry Interaction	Y	Y	Y	Y	Y	Rules-based scoring
Fill in the Blank—Uncon-strained Characters	N	Text Entry Interaction	Y	Y	Y	Y	Y	Rules-based scoring
	Y	Extended Text Interaction	Y	Y	Y	Y	Y	Handscoring
Fill in the Blank with Equation Editor	N	Text Entry Interaction	N	Y	Y	Y	Y	Rules-based scoring (Pearson uses proprietary systems to score)

PARCC Item Type	Valid Paper Type	QTI Interaction Type	ELA Item?	Math Item ?	Grade Band 3–5	Grade Band 6–8	Grade Band HS	Score Type
Category 3 TEI (basic)–QTI-defined interactions								
Drag and Drop	N	Order Interaction	Y	Y	Y	Y	Y	Rules-based scoring
	N	Match Interaction						Rules-based scoring
	N	Gap Match Interaction						Rules-based scoring
	N	Graphic Gap Match Interaction						Rules-based scoring
	N	Match Interaction						Rules-based scoring
Cloze	N	Choice Interaction	Y	Y	Y	Y	Y	Rules-based scoring
	N	Inline Choice Interaction						Rules-based scoring
Hot Spot	N	Hot Spot Interaction	Y	Y	Y	Y	Y	Rules-based scoring
	N	Position Object Interaction						Rules-based scoring
Non TE Response with TE Functionality	N	Dependent on TE Functionality	Y	Y	Y	Y	Y	Dependent on non-TE response
Inline Choice/ Drop Down	N	Inline Choice Interaction	Y	Y	Y	Y	Y	Rules-based scoring
PARCC Item Type	**Valid Paper Type**	**QTI Interaction Type**	**ELA Item?**	**Math Item ?**	**Grade Band 3–5**	**Grade Band 6–8**	**Grade Band HS**	**Score Type**
Category 4 TEI (innovative)–QTI custom interactions								
Line Graph	N	Custom Interaction	N	Y	Y	Y	Y	Rules-based scoring
	N	Custom Interaction						Rules-based scoring
Bar Graph	N	Custom Interaction	N	Y	Y	Y	Y	Rules-based scoring
Fraction Model	N	Custom Interaction	N	Y	Y	Y	Y	Rules-based scoring
Text Extraction/ Highlighting	N	Custom Interaction	Y	N	Y	Y	Y	Rules-based scoring

PARCC Item Type	Valid Paper Type	QTI Interaction Type	ELA Item?	Math Item ?	Grade Band 3–5	Grade Band 6–8	Grade Band HS	Score Type
Function Graph	N	Custom Interaction	N	Y	N	N	Y	Rules-based scoring
Interactive Number Line	N	Custom Interaction	N	Y	Y	Y	Y	Rules-based scoring
Zoom Number Line	N	Custom Interaction	N	Y	Y	Y	Y	Rules-based scoring
Geometric Trans-formations	N	Custom Interaction	N	Y	N	Y	Y	Rules-based scoring
Polygon Graph	N	Custom Interaction (class = tei-pointgraph_polygon)	N	Y	Y	Y	Y	Rules-based scoring
Solution Set Graphing	N	Custom Interaction	N	Y	Y	Y	Y	Rules-based scoring
Complex Drag and Drop	N	Custom Interaction	Y	Y	Y	Y	Y	Rules-based scoring
Complex Hot Spot	N	Custom Interaction	Y	Y	Y	Y	Y	Rules-based scoring
Composite Graphs	N	Custom Interaction	N	Y	N	Y	Y	Rules-based scoring

The integrity of the machine and human scoring processes are paramount to providing the Partnership rich, accurate field test data (to inform ongoing item selection for operational forms) and exacting operational test data for reporting results of the high-stakes tests.

Pearson will monitor all aspects of the scoring procedures, including key-based and rule-based machine scoring and handscoring for constructed response items and performance tasks. We possess established tools and processes to monitor scoring procedures throughout the entire data preparation, integration, scoring, analysis, and reporting processes.

The Pearson validation team will prepare test plans used throughout the scoring process. Test plan preparation will be organized around detailed specifications to be provided to PARCC at least three months prior to each administration.

At a minimum, the plan for PARCC will include:

- Raw score validation (e.g., score key validation; objective/strand/domain scoring, field test non-score; double-grid combinations; possible correct combination, if applicable; out-of-range / negative test cases)
- Derived scoring, if applicable (e.g., scaled score, performance level, and percentile score validation)
- Matching (e.g., validation of high-confidence criteria, low-confidence criteria, cross document, external or forced matching by customer; prior to and after data updates; extract file of matched and unmatched documents)
- Demographic update tests (e.g., verification of data extract against corresponding layout; valued values for updatable fields; invalid values for updatable / non-updatable fields; negative test for non-existing record or empty file)
- Aggregation, if applicable (e.g., tests of summary report data and field-level calculations; inclusion and exclusion criteria; minimum and maximum values for reporting categories; population subset confirmation; attemptedness rules).

These detailed plans are an integral component of the ongoing quality control processes we engage in throughout our program execution activities. Our well-developed procedures and policies promote thorough examination of our processing throughout these stages, and enable us to identify and resolve any issues that might arise.

We will score all student responses, as well as a sample of embedded field test items and a sample of external anchor items that do not count towards students' summative scores. Each representative sample will include 1,500 responses. In-depth explanation of the sampling procedures may be found in subsection Blueprint Sets of the PARCC Form Construction section under V.B.2 Summative Assessments.

Scoring procedures are detailed in the three sections that follow:

- Key-Based Scoring
- Rule-Based Scoring
- Handscoring

Capability to Meet PARCC Needs		
PARCC Requirement	**Understanding the Challenge**	**Feasibility, Scalability, and Completeness of Solution**
Key-Based Scoring (1)	It is essential that PARCC has the confidence that all key-based scoring data is accurate	Our quality engineers will use test plans and data sets to validate each key of the PARCC tests.
Rule-Based Scoring (2)	Rule-based scoring requires forward thinking and wide-ranging testing	With all scoring rules documented, our quality engineers will exercise and verify that common responses are considered. All rules-based items will be tested, reviewed, and updated as necessary.
Handscoring (3)	The handscoring program involves assigning more than 100 million scores in the first year of the program alone; hiring thousands of readers; and managing the program to the tight quality levels required PARCC while meeting schedule constraints required for reporting.	Our solution is based on the wisdom of four decades in the scoring business, past experience assigning more than 130 million scores in a single year and extensive experience scoring Common Core aligned items in Kentucky and elsewhere, and supporting item development and the tryout for PARCC. Our scale, scope, and experience uniquely position Pearson to manage the challenges and complexities of such a large-scale scoring program with foresight, skill, and a spirit of collaboration with PARCC stakeholders.

Meeting PARCC Program Needs. Pearson's scoring solution will address the scoring challenges and complexities of the PARCC assessment programs.

Key-Based Scoring

Prior to scoring student responses, it is essential that the scoring key information is accurate and any errors or discrepancies have been resolved. The item development process includes consistently and routinely updating scoring key information.

As part of our ISO 9000 standardized best practice test development process, we routinely perform a key review prior to test administration to verify that the scoring (answer) keys are correct for each form. Once the forms have been constructed and approved by PARCC for publication, an independent key review will be performed by experienced and trained content staff. The content staff will review each item and confirm that the key is correct.

If discrepancies are identified, a second senior content specialist or content manager will review the flagged item(s) and work with the item developers to resolve the issue. Our internal key review process will minimize the probability and risk that PARCC will identify errors in the scoring keys or will need to make additions to the scoring keys.

To complement the key review process, Pearson has developed a web-based application for the storage of test map data including the keys. The use of the application is transparent to the reviewer; however, it provides the key benefits of better version control (one source) and provides one location for item usage information. If there is a concern that must be addressed by our customers, program teams can log into the application and obtain needed information including comments and statistics on the item in question.

Another key review check takes place prior to the application being used to store test map data for the content specialist key review. Pearson's product support team runs a comparison of the multiple choice item keys in the customer test map against what is in the item code to check for discrepancies. All discrepancies are resolved internally and then will be presented to PARCC for approval.

When all key reviews have been completed prior to scoring, Pearson will provide PARCC the final keys for approval.

The test key information will then be handed off to PARCC for review and approval. We will work with PARCC on any outstanding discrepancies and/or errors and make the necessary updates or revisions in the item bank and resubmit the test key information for PARCC's final approval.

Rule-Based Scoring

PARCC will benefit from a reliable scoring method for technology-enhanced items that also allows timely delivery of results, particularly scoring for items designed to measure certain elusive, hard-to-measure performance standards. We are prepared to support this with our rule-based scoring solution.

Pearson employs a variety of scoring methods for educational assessments based on a number of factors, including item type and method of delivery. The previous section outlined the key-based scoring process that we will use to score PARCC's items through the use of scoring keys. This section outlines our process for scoring PARCC's technology-enhanced items based upon a set of previously determined scoring rubrics, a method also known as rule-based scoring. This method of scoring will be employed for multiple technology-enhanced item types, including gridded response, short constructed response, and other constrained constructed response items.

Scoring logics will differ by item type, and many item types will have multiple scoring logics. For instance, a line graph could be scored by either matching student-selected paired values for each plotted point to an answer key of paired values or inserting those paired values into an equation that defines a line or function.

Alternatively, a graphed line could be scored by slope: an exact value or a positive or negative slope. Similarly, a drag-and-drop item could be scored by requiring exact one-to-one pairings between draggers and drop zones (receptacles for those draggers), or each dragger could be associated with a value (e.g., the draggable quarter is associated with a value of 25) for evaluating the sum of the values for draggers moved to a particular area (e.g., provide a dollar in change).

Our staff will work closely with PARCC to first delineate the criteria for the scoring rubrics and then to adjust those rules based on student responses. The proposed scoring rubrics will be sent to PARCC for review, and if any additional changes are needed or new rules added, we will document and apply them.

During test construction, we will monitor and evaluate the scoring rubrics and update the correct answers/scoring rules in the item bank. We will submit the final scoring rules to PARCC for final approval.

Handscoring Overview

As stated in PARCC's grant application to the US Department of Education, "the strength of the assessment system rests on reliable, accurate, and efficient scoring" (PARCC, June 2010). During the Operational Assessment and continued field testing, accurate student scores will be critical to the overall success of the PARCC program. As a complex, multi-state endeavor, the PARCC program will require innovative approaches to scaling up handscoring, while driving unwavering consistency and reliability for all items, contents, and grades.

In order to meet these challenges, our scoring personnel have developed a solution for handscoring the Prose Constructed Response (PCR) items in ELA/Literacy (ELA/L) and constructed response (primarily Type II and Type III) mathematics items with a single goal: accurate, reliable scoring for PARCC and its stakeholders within the rigorous schedule constraints of the program.

Leaning on our experience with PARCC items through the item development process, our experience scoring large-scale, nationwide assessments, and our experience scoring in Partnership member states, our approach combines established methodologies with innovative strategies to increase scoring efficiency while maintaining rigorous standards of scoring quality.

To meet the needs of the PARCC program, Pearson offers the following:

- Valuable experience with PARCC items. As a current item development contractor and the prime contractor for Assessment Administration (item tryout and field test), our scoring experts have reviewed the items and rubrics as they were developed to provide input from a scoring perspective. Our team scored items during the research studies and the item tryout as part of the Assessment Administration contract, and we are currently preparing for field test scoring in spring of 2014. Our direct experience working with PARCC items will enhance consistency from field test to operational scoring, while reducing risk of implementation. Lessons learned from system set up, training, and scoring the field test will directly inform the handscoring specifications and planning for operational scoring and will greatly speed our scoring readiness.
- Broad experience scoring nationwide assessments (including scoring NAEP, college entrance exams, and ADP) combined with significant experience in Partnership member states. We currently score high-stakes assessments in Arizona, Colorado, Florida, Illinois, Kentucky, Maryland, and Ohio; we provide train-the-trainer and other services in New York; and we currently score the alternate assessment in the District of Columbia. Altogether, we have assigned nearly 41 million scores for PARCC states in 2013. Our deep talent pool of assigned project and content leadership have an average of 15 years' large-scale, high-stakes assessment experience and total of 55 cumulative years of scoring experience.
- Scale and capacity of distributed scoring with regional scoring hubs. Our solution combines the best of traditional scoring approaches with robust methods for taking this program to scale. Our distributed scoring model was built specifically to tap a nationwide pool of scoring resources, which makes it an ideal platform for consortia work. We are complementing our distributed scoring network with regional hubs where project and content management leads will oversee the work of handscoring leads and team leaders. These hubs will also act as call center support for distributed readers, and will include scoring as well. Because the Partnership Manager will play a vital role in guiding and monitoring reader training and scoring, the hubs will serve as a focal point for collaboration. Partnership representatives can be on-site at these regional hubs, interact with our staff, and observe training, scoring, and management.
- Next generation scoring platform. In 2010, we began to re-imagine and plan for the next generation of our scoring system to handle the volumes, item types, and complexity of the consortia assessments. The platform went live in the fall of 2013 for selected programs, and additional functionality is planned through 2014, including a next-generation content management and training platform. These innovative developments bolster our scoring engine, which already leads the industry for automation and quality management features.

- Automated scoring study and phase-in-plan. Automated scoring is predicated on high quality human scoring data. We have years of experience operationalizing human and automated scoring, and believe automated scoring provides a crucial opportunity for PARCC to drive scoring consistency and fast turnaround while mitigating the costs associated with human handscoring.
- Management rigor, stability, and consistency. All scoring will be managed under one set of procedures and one ISO-certified quality management system. Our approach eliminates unnecessary handoffs and variations in scoring approach, which can undermine consistency. No matter where a test item is taken, the student responses will be scored in the same way and to the same standards, reducing the risk that construct irrelevance may affect the outcomes. Pearson has the capacity and expertise to provide all the handscoring required by this program, and has been modeling handscoring volumes since PARCC first released its grant application proposal in June 2010 outlining preliminary scoring requirements. Further, should the program grow over time, beyond the scope outlined in the RFP, Pearson is well positioned to call on its collaborative, which includes handscoring experts, to support increasing volume.

The following figure shows the advantages and benefits of our scoring plan for field test and Operational Assessment scoring.

Pearson Handscoring Benefits Overview		
Feature	**Advantage**	**Benefit**
Next Generation Digital Scoring	Our patented digital scoring platform has evolved over 19 years. Scoring experts collaborate with engineers to build quality and automation tools into the system. Additionally, the same system supports both scanned paper assessments and online assessments.	Quality management is built directly into the system, enabling immediate action to be taken to enforce training and scoring rules and quality standards.
Nationwide Scoring Capacity	Reader quality and capacity with 24,000 experienced regional and distributed readers, and 100,000 screened candidates in our applicant database.	Ability to meet significant schedule and volume challenges while committing to accurate, reliable scores.
Robust Training	Comprehensive training for each item and task drives training rigor and scoring efficiency for the program. All readers will be trained and qualified prior to scoring, and will have practical experience scoring applying the rubrics to student responses through our pseudo scoring process.	Training will serve as the foundation for subsequent phases of the program, as well as help maintain accurate and consistent scoring across items.
More than 40 Years of Scoring Experience	Our best practices are built on more than 40 years of scoring experience, including scoring a wide range of complex performance items for the National Assessment of Educational Progress (NAEP) and other high-visibility state and national programs.	PARCC will be selecting a vendor with the depth of experience needed and the proven resource capacity to deliver this large, complex program.
ISO 9001:2008-Certified Quality Management System	Pearson's scoring services have operated under an ISO-certified quality management system for more than a decade. To our knowledge, Pearson is the first in the nation to achieve this status.	PARCC does not have to take Pearson's word for quality achievement; our scoring management and sites are routinely and rigorously audited as part of our ISO 9001:2008 certification. This certification is a framework for providing consistent processes across years, administrations, and scoring sites.
Scoring experience at scale—we are projected to assign over 80 million scores in 2014, and have assigned more than 130 million scores within a single year	Pearson has a depth of scoring talent, backed by scoring systems and processes to manage the work.	Pearson's experience on a range of scoring programs, including programs with tight scoring turnaround, translates to a lower risk, higher quality scoring program.
Expert scoring staff committed to scoring quality and accuracy	Pearson has selected highly qualified, seasoned scoring experts for this program. Pearson's lead scoring staff for the Operational Assessment has an average of 12 years with Pearson, and an average of 15 years teaching and assessment experience.	Pearson's staff is well-versed at managing all aspects of the scoring process, including managing multiple scoring programs and tight turnaround times for scoring services.

Scoring the PARCC Operational Assessment. Our scoring solutions will provide accurate, reliable scores for PARCC states.

The figure below summarizes our solution to meet PARCC's critical needs for reliable, scalable, and efficient handscoring of the approximately 90 million first and second scores that need to be assigned for the first traditional spring PBA operational administration over an approximate six-week scoring window. The table also provides a roadmap to our response and the individual sections that detail Pearson's solution to the RFP requirements.

Capability to Meet PARCC Needs		
PARCC Requirement	**Understanding the Challenge**	**Feasibility, Scalability, and Completeness of Solution**
Handscoring Overview (3)	The Partnership requires a Contractor to provide handscoring processes that are reliable and valid, as well as efficient in terms of time and expenditures for the Prose Constructed Response (PCR) items in ELA/Literacy and Type II and Type III items in mathematics.	Our handscoring solution combines proven processes and an ISO-certified quality management system with our next-generation scoring platform to deliver accurate and reliable scoring of all assessment items.
Rangefinding and Rangefinder Review Meetings (4)	PARCC requires a vendor that can manage an effective rangefinding plan with broad committee participation across member states for field testing as well as operational rangefinding in a hub city in a PARCC state three months prior to scoring each spring operational administration.	Rangefinding will be led by an expert team, with direct experience with PARCC items and experience leading 2014 field test scoring.
Produce Handscoring Materials (5)	Robust materials need to be created, reviewed, and approved in advance of scoring. Materials include scoring guides, training sets, qualifying sets, validity sets, group discussion sets, and recalibration sets.	Pearson will create a comprehensive set of handscoring materials to meet the needs of the PARCC assessments. Pearson staff oversees more than 300,000 hours of training delivery each year, and anticipate overseeing more than 500,000 hours of reader training in 2014.We understand the direct connection between the quality of the handscoring materials and the quality of the scoring outcomes.
Staffing for Handscoring (6)	PARCC requires a highly qualified team of scoring experts to oversee the entire scoring process, extending from rangefinding to scoring, with daily interaction with Partnership representatives during live scoring.	Pearson's highly qualified scoring staff has experience with high-stakes, large-volume assessments, including valuable experience with the PARCC 2014 field test. We have completed a comprehensive analysis of reader needs for the PARCC 2015 administrations, and have included this analysis in our proposal.
Team Leaders (7)	Teams will be led by trained, experienced scoring experts, responsible for groups of 10-12 readers and focusing on specific sets of performance tasks within each grade/subject.	Team leaders assigned to the PARCC project will have experience with other state or national assessments, and many will have worked on the PARCC field test in 2014.
Recruit and Hire Readers (8)	PARCC readers must be carefully selected and qualified to score student responses.	Our human resources staff will recruit and hire readers who meet strict qualifications to score the PARCC assessments, from our pool of experienced readers and our database of 100,000 screened applicants.

Capability to Meet PARCC Needs		
PARCC Requirement	**Understanding the Challenge**	**Feasibility, Scalability, and Completeness of Solution**
Training and Qualifying of Readers (9)	PARCC readers will complete rigorous training in preparation for scoring, and must meet PARCC qualification standards.	Our online modules deliver a robust training regimen that will align PARCC readers with Partnership Management standards for scoring. Training and qualifying rules are automatically enforced by the system.
Handscoring Reports (10)	Partnership Management requires cumulative and daily handscoring reports on reader performance and quality.	Our next-generation scoring platform allows for a range of reports on reader performance and quality, which meet PARCC requirements and are available online for Partnership Manager staff.
Scoring Student Responses (11)	Student responses will be scored with 20 percent second scoring and will undergo resolution scoring as necessary. Discussion sets will be used to maintain reader alignment with scoring standards, and alert systems will be in place to manage student responses that need intervention.	Pearson will configure system scoring rules to PARCC requirements. System set up will be greatly facilitated by experience scoring the 2014 PARCC field test. The system will automatically route responses for scoring, and will support reader intervention and alerts. Further, the new system architecture was specifically designed to support the scale of consortia work.
Monitor and Maintain Handscoring Quality (12)	PARCC requires high levels of scoring accuracy for scoring student responses, as well as measures to quickly address scoring issues.	Our scoring system is designed for constant monitoring of handscoring quality, allowing scoring management staff to intervene quickly and effectively. Our scoring system offers a robust suite of automation tools, which take action to intervene (messages, additional training papers, lockouts), so readers performing below expectations receive swift intervention to prevent reader drift.
Special Scoring Requirements for NY (13)	Scoring solutions will include options to meet special scoring requirements for New York State, including teacher scoring of student responses.	Pearson is familiar with the special circumstances of New York scoring, having worked with New York on previous assessment programs. Our solution provides three options for New York scoring.
Deliverables V.B.2.K	PARCC management requires timely, high quality scoring deliverables that are approval-ready.	Our team is well versed in the specific deliverables required for this program, and will deliver high quality results in a timely manner.

Meeting PARCC Scoring Needs. Pearson offers PARCC significant capability for each major scoring requirement in the RFP.

Scoring System

Delivering a Comprehensive Scoring Solution for PARCC

In Year 1 of the PARCC handscoring program, our readers will assign nearly 90 million operational scores and nearly 2 million field test scores. For a program of this scale and complexity it is imperative that the scoring system, used by trained readers to score PCRs and mathematics items, meet the performance, configuration, functionality, and throughput demands of the program.

Pearson scoring experts have evaluated the number of open-ended ELA/literacy and mathematics items by grade to provide recommendations for training, scoring and monitoring reader quality.

Our computer-based distributed scoring solution blends scoring technology with extensive training and scoring expertise. With our solution, PARCC can expect scoring efficiency and accuracy, and can increase participation opportunities for readers throughout PARCC states, or outside of PARCC states, depending on state preferences. Our distributed scoring platform was built specifically to reduce barriers to participation and tap a nationwide pool of scoring experts, making it an ideal platform for multi-state assessments.

Pearson's considerable experience in scoring for a number of PARCC states, as illustrated below, will prove valuable in implementing a multi-state program.

Total Scores Assigned by Pearson in PARCC States		
PARCC States	**Arizona, Colorado, Washington, DC, Florida, Illinois, Kentucky, Maryland, New Jersey, New York, and Ohio**	
	2012	**2013**
Total	42,476,000	40,926,000

Scoring Experience in PARCC States. Pearson currently provides scoring solutions for assessments in key PARCC states.

Next Generation Scoring System

Our digital scoring platform is currently being configured for and will support scoring of the 2014 PARCC field test. Consequently, all of the item and rubric types that will eventually be scored as part of the 2014–15 Operational Assessment will already have been configured in our system, fostering ease of implementation and reduced risk.

As the pioneer of computer-based scoring within the assessment industry, Pearson owns a comprehensive series of patents covering components of digital scoring for performance assessments. Our patented digital scoring platform has evolved over 19 years, with scoring experts and engineers collaborating to integrate quality and automation tools into the platform. The system, in use in the US and internationally, has exceptional capacity, and has been used to assign as many as 130 million scores in a single year for US programs alone.

Unlike other vendors' systems, Pearson's computer-based scoring system is lean by design – work is pulled by the reader versus being pushed into their queue, where it might sit idle during non-scoring hours. Our system has efficient response routing features that eliminate processing lag times while offering robust automation tools to support quality management.

With our digital scoring system, the scoring process is streamlined and controlled:

- Responses are scanned from original test books and image clips distributed to qualified and trained readers, or online responses are uploaded into the system.
- The digital scoring system automatically routes responses requiring second scores or resolution reads to qualified personnel.
- The system automatically prioritizes responses that need to be scored first, such as responses for equating batches.
- All scores assigned to responses are automatically captured and available for review.
- Digital scoring integrates multiple processes (routing work, scoring responses, monitoring quality, and tracking progress and workflow) into a single, efficient, user-friendly system.
- Pearson's platform routes work between human and artificial intelligence scoring; based on outcomes of our efficacy study, our solution reduces the risks and fosters the transition to human-automated scoring applications.

Our next-generation, web-based scoring system is designed to meet the growing demand for scalable, internationalized, and regional and distributed performance scoring. Our system is built on the latest technologies and uses the scalable hardware environment of the cloud. The following table highlights key features of our system:

Next-Generation Scoring System Features
Accessible from any Internet-connected computer so readers can work in regional or distributed models.
Access is granted only to authorized users, and then only to features, data, and uploaded content appropriate to the user's role.
Student content is anonymous. Readers do not know which content belongs to which student and therefore must score based only on the content itself.
Content is captured and delivered to readers electronically. There are no packets of papers to ship, track, or distribute to readers and nothing to retrieve later.
A scoring guide in the item scoring view provides visual feedback for score point selections and support for a variety of response media, including audio, video, and numerous image formats.
Increased accuracy, because team leaders have the tools they need to discover and mitigate problems before they affect item performance. These include quality analytical tools and automated quality monitoring. Score quality can be measured in real time.
Reports aggregate the most salient performance and quality information and differentiate it in a dashboard-like format on the team leader's home screen so it can be easily compared, consumed, and when necessary, acted on.
Reports also act as a portal to drill down to in-depth information on specific performance areas and provide shortcuts directly into the scoring tasks and queues. This increases efficiency, improving the visibility of quality metrics, and maximizes the speed at which team leaders can moderate scoring issues.
Easy communication between readers and administrators allows for helpful, timely feedback which increases the precision of scoring.
Manages scanned paper responses and text-based responses using the same platform, for ease and efficiency of scoring

Scoring System Features. PARCC will receive accurate, timely results with the comprehensive features of the Pearson computer-based scoring system.

Ongoing Investments in the Digital Scoring System

Pearson has more than a decade of experience refining our platform and developing next-generation capabilities, particularly in the area of reader management and quality control. Enhanced integration with other Pearson products provides expanded training and content management, improves flexibility to using automated scoring engines, and enables wider use of metadata to drive scoring relationships among various modes of scoring (for example, human and automated, professional readers and teacher-readers, etc.).

Based on feedback from hundreds of users, the new interface design improves user experience and drives efficiencies for managing scoring activity. Examples of the user interface can be found in the Scoring Student Responses section below.

Features of our next generation scoring system of particular relevance to consortia work are highlighted in the figure below.

Pearson's Next Generation Scoring System: Flexibility, Scalability, and Ease of Use		
Features	**Advantages**	**Benefits**
Learning Management System	The learning management system has a user-friendly interface and can be used in concert with operational scoring or separately, to provide educators valuable access to scoring training materials for professional development. The LMS features many built-in help tools and opportunities for readers to collaborate with scoring experts.	The new LMS is easy to use and will promote ease of training for the thousands of readers assigned to the PARCC project.
Content Management System	The new content management system will allow for an online workflow and approval system for scoring training materials and annotations.	The many stakeholders involved in the training review and approval process will have an easy-to-use interface for training set review and approval.
Scoring Broker	This module within the overall system architecture will enable routing for human, automated, and outlier scoring, based on pre-defined parameters. The scoring broker will also support special scoring rules, including routing work to New York teachers if desired.	Items and tasks will be efficiently routed for scoring by mode, without delays caused by batch processing or manually loading responses to disparate systems.
Scalability	Pearson first started modeling Common Core scoring requirements in 2009-2010. We used this modeling data to inform the architecture and scalability of the new system.	The system is scalable to be able to support scoring numbers, with the ability to rapidly add capacity as needed. The system was designed to support volumes articulated in the 2010 PARCC grant proposal, which exceed the volumes in the current RFP and allow for growth of the program over time.

Digital Scoring Benefits. The next generation of our scoring system was designed specially with PARCC requirements in mind, based on an analysis of PARCC's 2010 grant proposal.

Trained Professional Readers

Qualified Readers for the PARCC Assessments

Readers will review and evaluate each response according to the rubric and anchor responses to assign appropriate scores. We will select readers based on their academic and/or professional backgrounds and scoring experience and require that all readers possess at least a four-year college degree. Our professional readers are highly qualified, capable, and intensively trained and monitored, using validity, reliability, and other measures. Readers are important contributors to our achieving and maintaining a high degree of consistency and reliability in scoring responses.

Pearson currently has a pool of more than 24,000 experienced regional and distributed readers, and a screened applicant pool of more than 100,000 readers nationwide. This substantial pool of talent will allow us to carefully select and screen readers based on the requirements of each program.

Further information about our recruiting and hiring procedures can be found in the Recruit and Hire Readers section. Pearson's capacity to meet the training and qualification needs of the PARCC program is outlined in the Training and Qualification section. Full details on the system tools for monitoring reader performance and quality can be found in the Monitor and Maintain Handscoring Quality section.

Scoring Rules

Scoring Rules for PARCC Assessments

In Year one, all of the ELA/L PCRs will receive a first score by a human reader. Ten percent of the student responses will also receive a second score. Second scores will be assigned by a reader for paper-based responses and assigned using artificial intelligence (AI) for online responses.

In Year two, two-thirds of the online ELA/L PCR items per grade will receive the first score using AI scoring, with a 10 percent second score done by a reader. The remaining one-third of the online PCRs per grade will have the first score applied by a reader with 10% second score assigned by the AI scoring engine.

In Years three and four, all online ELA/L PCRs will receive their first score from AI scoring with 10% scoring done by readers. The table below summarizes the human and automated scoring plan for online ELA/L responses.

Year	% of ELA/L items	1st Score (100%)	2nd Score (10%)	Resolution
2015*	100%	Human	Automated	Human
2016	67% 33%	Automated Human	Human Automated	Human Human
2017	100%	Automated	Human	Human

Across all years, any responses scored by the automated scoring engine outside predetermined confidence levels established for automated scoring will be scored by a human reader. These responses, also known as "outliers," are often atypical responses (for example, particularly short). We have planned for up to 5% of the responses to be routed for human scoring as outliers.

In all years (one through four), all Mathematics constructed response items will receive a first score by a human reader. Ten percent of the student responses will also receive a second score whether the response is online or via paper.

Across the years, if the first and second score are nonadjacent, a third human and sometimes a fourth human reader shall be used. These scoring rules apply across scoring modes (human or automated) and test delivery types (paper or online). Note that the first score will be used to calculate the student's final score.

The digital scoring system will automatically and randomly distribute responses for second scoring; readers will have no indication whether a response has been scored previously or knowledge of the previous score.

Scoring Sites

Distributed and Regional Scoring

In order to meet the capacity and volume needs of such a complex, large-scale assessment program, we propose a distributed scoring model for handscoring, supported by regional scoring hubs. The regional hubs will house project and content staff, call center staff (content, technical, and human resources), supervisors and readers (to augment the distributed base). The regional hubs will also allow PARCC representatives direct oversight of the scoring process. This solution combines the best of traditional scoring approaches in regional centers with the robust methods and technology needed to accommodate the scale of the PARCC program.

Distributed Scoring

Pearson began using distributed scoring in 2002, and we continue to refine and update our scoring system. The system incorporates several innovative components, including:

- Online training and qualification are comprehensive and item specific.
- Extended-hours performance scoring support centers in our regional scoring hubs to provide quality monitoring, feedback, user support, and technical help using a wide range of industry leading tools.
- Comprehensive scoring and monitoring tools, including backreading, calibration, and reporting along with advanced automation features built directly in our scoring system.
- Pearson has mature, repeatable processes that we have refined over the past decade for managing distributed projects. The distributed scoring methodology is used to score high-stakes state assessments used for federal and state accountability as well as national assessments, including college entrance exams. In 2013, of Pearson's 37 scoring programs that require performance scoring services, over 50 percent use distributed scoring, and more are adopting this model every year.

- Our distributed readers come from a dedicated and highly credentialed pool, with more than 60 percent of our 2013 active readers holding a master's degree. By the end of 2013, over 50 percent of total responses scored by Pearson will be scored by our distributed workforce.

Distributed scoring offers the following benefits:

- **Flexibility in scheduling and staffing.** Flexible scoring hours allow more readers to contribute to a project, shortening scoring schedules. With a wider and more diverse pool of readers comes the potential to better match readers with projects.
- **Involving educators in the scoring process**, key to a solution for New York. Distributed scoring provides an opportunity for teachers from across all parts of New York to participate, in addition to scoring through facilities provided by the BOCES.
- **Risk mitigation.** With distributed scoring, we are not vulnerable to operational disruptions related to weather or other events, nor is reader recruitment narrowly tied to specific regional labor markets.

Supporting Research

The effectiveness of the proposed distributed scoring model is backed up by academic research, including four formal research studies conducted from 2007 to 2009. The initial studies compared the quality of practice scoring, qualification rates, and scoring of operational responses achieved in online training and stand-up training. Later studies also included the two methods of training and the results of distributed scoring and regional scoring. Collectively, these studies demonstrate comparable quality and effectiveness between regional and distributed scoring methods.

The published results of the studies above indicate:

- Perfect agreement across distribution of final essay scores for primary testers
- Near perfect agreement across distribution of final essay scores for retesters
- 95–98 percent consistency of student classification at the total test–level between regional and distributed scoring
- 95–98 percent consistency of student classification at the total test–level between study and operational[1]

Performance Scoring Support for Distributed Scoring

To meet the quality standards of PARCC, we will manage the scoring through our performance scoring support center, where staff will provide quality monitoring and feedback, human resources assistance, and technical support.

[1] Wolfe, E.W., Matthews, S., & Vickers, D. (2010). The Effectiveness and Efficiency of Distributed Online, Regional Online, and Regional Face-to-Face Training for Writing Assessment Raters. *Journal of Technology, Learning, and Assessment, 10*(1)

Our performance scoring support centers, located in our PARCC regional hubs, will be the focal point for project leadership and monitoring. Team leaders at the performance scoring support centers will monitor scoring quality, respond to scoring questions, and offer content-related support to distributed readers.

> *Our performance scoring support centers for PARCC Assessments will operate under the leadership of our project management team. It will be staffed with content specialists and team leaders with substantial personal experience and knowledge of scoring assessments.*

In addition to training, our performance scoring support staff are well versed in supporting distributed readers. In 2013, our call center responded to nearly 100,000 calls, emails, and chat sessions from distributed readers.

Reader Communication

Reader communication promotes scoring consistency and helps proactively address reader questions and concerns before issues occur. Team leaders are the first point of contact for reader questions. If needed, questions can be escalated to handscoring leads or content specialists for resolution.

Project scoring website. A secure website will be created specifically for the PARCC handscoring program. The site will house reader-facing quality management plans, project scoring dates, contact information, scoring materials for downloading (once approved by PARCC), and content FAQs. Scoring support staff will post news to the project's secure and dedicated website informing readers of project updates, including critical schedule information.

Phone system, email, and chat. The PARCC handscoring program will have a dedicated toll-free phone number for readers to use when they have questions. Support will consist of three groups who take inbound contacts: scoring content representatives, human resources, and general representatives who also field technical questions. Calls will be directed to a group based on the caller's selection from a simple interactive call menu system.

The support center will also provide email support through a customized PARCC handscoring support email address. Additionally, our PARCC support team will provide real-time employee support via chat, allowing readers the flexibility to interact with agents online. Note: To maintain stringent security standards, no content or questions concerning content will be exchanged via email or chat.

The staff in the performance scoring support centers will monitor inbound calls and email messages so that all questions are answered in a timely manner or directed to the appropriate group. All contacts will be tracked in a software program called *Service Manager*, allowing call center staff to review all inquiries posed by the reader and responses given by other staff. This documentation promotes consistency and also provides an audit trail for quality management purposes.

Secure messaging. Within our digital scoring system, we offer secure messaging, which can be used to message an individual reader or an entire group of readers. Readers will be able to send student responses into a queue for review and ask a team leader questions about the response. Team leaders will be able to give direct feedback to readers while allowing them to review the response. PARCC team leaders will also use this tool while backreading.

With these three different platforms (secure website, call center, and secure messaging) we efficiently and effectively answer specific reader questions individually and without unnecessarily interrupting other readers.

Distributed Scoring Quality Management

All project-related quality management measures will be outlined in a Quality Management Plan (QMP), which will be part of the handscoring specifications. The QMP states the specific project requirements and how actions regarding readers should be taken. All project-related actions regarding readers will be tracked so that all supervisory staff members are aware of the disposition of every reader.

Quality management measures are established by the project management staff and are administered by the team leads. These processes include, but are not limited to, qualification agreement rates, calibration results, inter-rater reliability (IRR), validity, and scoring rate. If quality issues are found, the following tools in our scoring system can be used to aid in improving reader accuracy: validity review, calibration scoring, backreading, messaging, review, and resolution scoring/adjudication.

Automated tasks that are managed by our scoring system include, but are not limited to, validity checks, scoring rate checks, and automated lockouts based on scoring hours. These tasks will be set up prior to the start of scoring and can be tailored to specific requirements. Additionally, these tools can be used to measure a reader's accuracy and, if needed, lock out the reader if he or she is not meeting project standards. Full details on quality management practices can be found in the Monitor and Maintain Handscoring Quality section below.

Distributed Scoring Security

Pearson's scoring system provides secure transmission of data at login and during active sessions through the use of industry-standard Secure HTTP (HTTPS) and Secure Socket Layer (SSL) technology. We also follow industry standard access, password, and user identification protocols when authenticating users to our digital scoring system.

Readers will be trained on security protocols and will be required to sign a confidentiality agreement. All student responses will be distributed anonymously, with no way for a reader to link a student's demographic information to a response.

For security reasons, readers may not use shared office computers or work from institutional or public computer labs. Pearson has taken measures to prevent distributed readers from printing student responses in order to guard against unauthorized disclosure of test content.

Unparalleled Scoring Capacity—Regional Sites

In order to accommodate the volume of student responses anticipated for the PARCC Operational Assessments, distributed scoring will be managed from and augmented by regional scoring hubs. Pearson offers decades of experience managing regional locations. In the past 10 years alone, we have set up more than 120 permanent and temporary scoring centers, ranging from small sites to support teacher scoring or professional development workshops to large, 500-seat operations. Pearson anticipates having well over 333,500 square footage of regional scoring capacity in 2014, evidence of our capability to manage both regional and distributed scoring at scale.

For the PARCC handscoring program, our staff will carefully plan out the regional hubs, using lessons learned from the 2014 PARCC field test, which will be scored in sites. Our human resources and facility staffs will review performance at existing sites and research potential new scoring sites with the understanding that our regional hubs should be located in a city that provides a pool of well-educated individuals from which to recruit and fosters ease of travel for Partnership Management staff.

By establishing scoring locations across the country and in cities that have high numbers of degreed professionals, we not only increase the number of trained readers available for our projects, but also enrich the diversity and professional makeup of our reader and staff pool.

The scoring centers in our nationwide network, as shown in the following figure, operate under our ISO-certified quality management system and increase our flexibility to complete scoring in an efficient and timely manner.

Planned Pearson Scoring Centers for 2014				
Scoring Site	**In Metro Area Since**	**Capacity**	**Approximate Square Footage**	**PARCC State**
Austin, TX	1989	480	25,000	
Charlotte, NC	2009	250	15,000	
Chicago, IL (Lombard)	2010	300	25,000	X

Planned Pearson Scoring Centers for 2014				
Scoring Site	**In Metro Area Since**	**Capacity**	**Approximate Square Footage**	**PARCC State**
Colorado Springs, CO	2014	200	8,000	X
Columbus, OH (Westerville)	2005	300	16,000	X
Hadley, MA	2006	100	4,000	X
Houston, TX	2008	200	8,000	
Iowa City, IA	1990	100	3,000	
Jacksonville, FL	2000	300	20,000	X
Lansing, MI	1999	350	21,000	
Lexington, KY	2014	200	8,000	X
Mesa, AZ	2001	300	40,000	X
Nashville, TN	2014	200	8,000	X
San Antonio, TX	1988	600	30,000	
San Juan, PR	2010	110	10,000	
Seattle, WA (Kent)	2003	300	27,000	
Tucson, AZ	1998	300	20,000	X
Tulsa, OK	2009	200	8,000	
Virginia Beach, VA	2002	300	37,500	
Total Square Footage			333,500	

Pearson Scoring Capabilities. Pearson has extensive regional capacity, unified through a common leadership vision and repeatable processes, and backed by the rigor of an ISO-certified quality management system. In 2014, Pearson will have more than 5,090 seats for one shift alone across more than 333,500 square feet of facilities.

For the PARCC Assessments, we will select locations for our regional scoring hubs, where the PARCC representatives can oversee scoring and Pearson leadership will train and monitor staff. These sites will also house readers and supervisors, so that some scoring for each content area is regionally based. This will allow us to promote readers into supervisory roles, for example, and allow us to record reader discussions for use in online discussion sets.

Security at All Pearson Scoring Locations

Pearson prioritizes an environment that promotes the security of test items, student responses, data, and employees throughout the project. We are able to accommodate the needs of our customers while employing strict safeguards for security at our sites:

- All locations have secure, monitored entrances and exits. Access is restricted to full-time, part-time, and authorized visitors.
- All staff must wear Pearson identification badges at all times in the scoring facility.
- All authorized visitors must sign in, be escorted at all times, and wear identifying badges.
- All scoring personnel are trained on security protocols and must sign a nondisclosure and confidentiality form in which they agree to not use or divulge any information concerning test processes, scoring guides, or individual student responses.
- Secure materials are accounted for daily and maintained in locked storage.
- Our readers are not able to print from their workstations.
- Use of recording or photographic equipment is not allowed in the scoring area.
- Test materials shall not leave the facility during the project without the permission of a person or persons designated by Partnership Management.

Handscoring Specifications

A program of the size, scope, and magnitude of the PARCC operational and field test administrations must be governed by carefully constructed, customer-approved specifications outlining the scoring design, schedule, metrics, meetings, and other key facets of the program. Creating comprehensive hand-scoring specifications is crucial to obtaining accurate and consistent scores and is an important part of our ISO 9001:2008 registered quality management system.

Pearson will create hand-scoring specifications during the planning phase and deliver to the Partnership Manager at least four months prior to testing. The handscoring specifications will be a detailed guide to conducting handscoring and will be used by our scoring managers, scoring content specialists, handscoring leads, and PARCC.

Pearson will provide the handscoring specifications, including all of the information stated in Section V.B.2.K.3.e.1-18 of the RFP and summarized below:

- Handscoring schedules
- Site requirements for the regional hubs (including security and access)
- Scoring design (including expected number of reads; groupings of items assigned to readers; assignment of first and second reads; and allocation of scoring by regional and distributed)
- Personnel (including requirements for handscoring leads, team leaders, readers; number of personnel; and security agreements)
- Rangefinding meetings (including staffing and meeting logistics and procedures)
- Rangefinder review meetings (including staffing and meeting logistics and procedures)

- Descriptions of training and qualifying materials, scores, and annotations
- Descriptions of validity sets (including number and type of validity responses
- Reader degree verification procedures
- Training procedures
- Qualifying standards
- Handscoring reports (including list and description of reports and distribution)
- Scoring process (including administration of validity sets; rules for determining and assigning resolution reads; monitoring and retraining; read-behind guidelines; and use of daily and cumulative reader reports
- Processing requirements (including configuration and set up parameters for the scoring system)

Pearson will also provide detailed documentation of the scoring process:

- Electronic copies of annotated training materials for PARCC/PRC distribution to districts.
- On-line access to all handscoring systems and reports to support (during live scoring daily Handscoring status calls will also be held)
- Handscoring report at the conclusion of scoring, including:
  - Item or task ID and form information
  - Total number of reads
  - Inter-Rater Reliability statistics
  - Validity Statistics
  - Score point distribution for scores from handscoring and the mean score
  - Final quality-assurance monitoring reports from the handscoring system
  - Scoring notes from handscoring
  - Summary of rangefinding, including staff, meetings, etc.

The statistical data will be provided electronically in a format agreed upon by the Partnership Manager and Pearson.

Our staff have analyzed the test design, item types, and rubrics to be scored under this program. The following figures summarize the scoring scope for the PARCC field test and Operational Assessments each year of the contract, which will be covered in detail in the handscoring specifications. Note: Volumes in these tables represent the combined total of paper and online responses. Also, scoring for field test items administered in the last year will be scored as part of a contract extension or by the next contractor. Additionally, state participation levels and final student volumes will affect the number of responses to score in these tables.

2014-15 PARCC Field Test - Regional Scoring Model									
English Language Arts / Literacy									
Grade	3	4	5	6	7	8	9	10	11
Responses per Item	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500
Unique Items	42	42	42	42	42	42	48	48	42
PCR Items / Student (5 pt)	3	3	3	3	3	3	3	3	3
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring*	10%	10%	10%	10%	10%	10%	10%	10%	10%

2014-15 PARCC Field Test - Regional Scoring Model						
Mathematics						
Grade	3	4	5	6	7	8
Responses per Item	1,500	1,500	1,500	1,500	1,500	1,500
Unique Items	61	61	61	61	61	61
Items per Student	7	7	7	7	7	7
Testing method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second scoring	10%	10%	10%	10%	10%	10%

High School Mathematics						
Course	Algebra I	Algebra II	Geometry	Int Math I	Int Math II	Int Math III
Responses per Item	1,500	1,500	1,500	1,500	1,500	1,500
Unique Items	83	83	88	56	56	69
Items per Student	8	8	10	8	8	10
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%

2015-16 PARCC Field Test - Regional Scoring Model									
English Language Arts / Literacy									
Grade	3	4	5	6	7	8	9	10	11
Student Volume	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500
Unique Items	42	42	42	42	42	42	48	48	42
PCR Items / Student (5 pt)	3	3	3	3	3	3	3	3	3
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring *	10%	10%	10%	10%	10%	10%	10%	10%	10%

*Ten percent of paper based responses will receive a second score. Five hundred online responses will receive a second human for the purpose of generating sufficient double-scored cases to train the automated scoring engine. This applies to online responses only.

2015-16 PARCC Field Test - Regional Scoring Model						
Mathematics						
Grade	3	4	5	6	7	8
Student Volume	1,500	1,500	1,500	1,500	1,500	1,500
Unique Items	61	61	61	61	61	61
Items per Student	7	7	7	7	7	7
Testing method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%
High School Mathematics						
Course	Algebra I	Algebra II	Geometry	Int Math I	Int Math II	Int Math III
Student Volume	1,500	1,500	1,500	1,500	1,500	1,500
Unique Items	83	83	88	46	46	56
Items per Student	8	8	10	8	8	10
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring %	10%	10%	10%	10%	10%	10%

2016-17 PARCC Field Test - Regional Scoring Model									
English Language Arts / Literacy									
Grade	3	4	5	6	7	8	9	10	11
Student Volume	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500
Unique Items	36	36	36	36	36	36	42	42	36
PCR Items / Student (5 pt)	3	3	3	3	3	3	3	3	3
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring*	10%	10%	10%	10%	10%	10%	10%	10%	10%

*Ten percent of paper based responses will receive a second score. Five hundred online responses will receive a second human for the purpose of generating sufficient double-scored cases to train the automated scoring engine. This applies to online responses only.

2016-17 PARCC Field Test - Regional Scoring Model						
Mathematics						
Grade	3	4	5	6	7	8
Student Volume	1,500	1,500	1,500	1,500	1,500	1,500
Unique Items	50	50	50	50	50	50
Items per Student	7	7	7	7	7	7
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%
High School Mathematics						
Course	Algebra I	Algebra II	Geometry	Int Math I	Int Math II	Int Math III
Student Volume	1,500	1,500	1,500	1,500	1,500	1,500
Unique Items	69	69	72	28	28	35
Items per Student	8	8	10	8	8	10
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%

2015-16 PARCC									
English Language Arts / Literacy									
Grade	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Student Volume	1,075,206	1,085,252	1,080,828	1,071,556	1,073,060	1,060,066	552,954	522,308	478,772
Unique Operational Items	15	15	15	15	15	15	15	15	15
PCR Items / Student (5 pt)	3	3	3	3	3	3	3	3	3
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%	10%	10%	10%
Team Lead to Reader Ratio	1:12	1:12	1:12	1:12	1:12	1:12	1:12	1:12	1:12

2015-16 PARCC						
Mathematics						
Grade	3	4	5	6	7	8
Student Volume	1,075,206	1,085,252	1,080,828	1,071,556	1,073,060	1,060,066
Unique Operational Items	36	36	36	36	36	36
Items per Student	7	7	7	7	7	7
Testing method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second scoring	10%	10%	10%	10%	10%	10%
Team Lead to Reader Ratio	1:12	1:12	1:12	1:12	1:12	1:12
High School Mathematics						
Course	Algebra I	Algebra II	Geometry	Int Math I	Int Math II	Int Math III
Student Volume	414,714	359,076	391,732	138,238	130,576	119,694
Unique Operational Items	43	54	43	6	6	7
Items per Student	8	8	10	8	8	10
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%
Team Lead to Reader Ratio	1:12	1:12	1:12	1:12	1:12	1:12

2016-17 PARCC									
English Language Arts / Literacy									
Grade	3	4	5	6	7	8	9	10	11
Student Volume	1,075,206	1,085,252	1,080,828	1,071,556	1,073,060	1,060,066	552,954	522,308	478,772
Unique Operational Items	15	15	15	15	15	15	15	15	15
PCR Items / Student (5 pt)	3	3	3	3	3	3	3	3	3
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%	10%	10%	10%
Team Lead to Reader Ratio	1:12	1:12	1:12	1:12	1:12	1:12	1:12	1:12	1:12

2016-17 PARCC						
Mathematics						
Grade	**3**	**4**	**5**	**6**	**7**	**8**
Student Volume	1,075,206	1,085,252	1,080,828	1,071,556	1,073,060	1,060,066
Unique Operational Items	36	36	36	36	36	36
Items per Student	7	7	7	7	7	7
Testing method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second scoring	10%	10%	10%	10%	10%	10%
Team Lead to Reader Ratio	1:12	1:12	1:12	1:12	1:12	1:12
High School Mathematics						
Course	**Algebra I**	**Algebra II**	**Geometry**	**Int Math I**	**Int Math II**	**Int Math III**
Student Volume	414,714	359,076	391,732	138,238	130,576	119,694
Unique Operational Items	43	54	43	6	6	7
Items per Student	8	8	10	8	8	10
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%
Team Lead to Reader Ratio	1:12	1:12	1:12	1:12	1:12	1:12

2017-18 PARCC									
English Language Arts / Literacy									
Grade	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Student Volume	1,075,206	1,085,252	1,080,828	1,071,556	1,073,060	1,060,066	552,954	522,308	478,772
Unique Operational Items	15	15	15	15	15	15	15	15	15
PCR Items / Student (5 pt)	3	3	3	3	3	3	3	3	3
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%	10%	10%	10%
Team Lead to Reader Ratio	1:12	1:12	1:12	1:12	1:12	1:12	1:12	1:12	1:12

2017-18 PARCC						
Mathematics						
Grade	3	4	5	6	7	8
Student Volume	1,075,206	1,085,252	1,080,828	1,071,556	1,073,060	1,060,066
Unique Operational Items	36	36	36	36	36	36
Items per Student	7	7	7	7	7	7
Testing method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second scoring	10%	10%	10%	10%	10%	10%
Team Lead to Reader Ratio	1:12	1:12	1:12	1:12	1:12	1:12
High School Mathematics						
Course	Algebra I	Algebra II	Geometry	Int Math I	Int Math II	Int Math III
Student Volume	414,714	359,076	391,732	138,238	130,576	119,694
Unique Operational Items	43	54	43	6	6	7
Items per Student	8	8	10	8	8	10
Testing Method	online / paper	online / paper	online / paper	online / paper	online / paper	online / paper
Second Scoring	10%	10%	10%	10%	10%	10%
Team Lead to Reader Ratio	1:12	1:12	1:12	1:12	1:12	1:12

The high-stakes nature of the PARCC Operational Assessments demands that handscoring meet stringent levels of scoring accuracy while adhering to scoring schedules. Since Pearson began providing scoring services, we have served increasingly large and complex assessments for a diverse array of programs. To maintain our leadership in the scoring industry, we continually update and refine our scoring systems and processes. Our experience with high-stakes, high-volume assessments allows us to be prepared to the volumes and demanding schedules anticipated for the PARCC assessments. Details on our scoring management and quality monitoring processes can be found in the Scoring Studies Responses section and the Monitor and Maintain Scoring Quality section below.

Pearson understands that security of test items and student responses is vital to a successful assessment program, and prioritizes an environment that promotes the security of test items, student responses, data, and employees. Information on our security measures for both distributed scoring and scoring at our regional scoring hubs can be found in the Scoring Locations section.

The Partnership Manager will be crucial to planning and implementing the assessment program. Pearson expects that the Partnership Manager will be integrally involved throughout the rangefinding and rangefinder review meetings, training of readers, and scoring sessions, and will be on site during critical periods of the program. On or off site, the Partnership

Manager will have access to Handscoring reports and will receive consistent communication from Pearson scoring management staff, including daily status calls during live scoring.

Rangefinding and Rangefinder Review Meetings

Field Test Rangefinding

Field testing new items will be vital to maintaining the quality of the PARCC assessment program. Field test rigor is especially important to support high quality data for calibrating the automated scoring engine.

In collaboration with PARCC, Pearson will select a range of student responses to field test items. These responses will be presented in field test rangefinding meetings. Documented outcomes of these rangefinding meetings will be used to create training and qualifying sets to train readers on new items.

As requested in the RFP, we will present approximately half of the field test items in rangefinding meetings in the spring after the conclusion of the operational test; the remaining field test items will be presented to rangefinding committees in the fall. Upon the completion of rangefinding in both the spring and fall sessions, Pearson staff will create and organize training, qualifying, and validity materials for training readers.

Field test rangefinding and scoring will be completed following in 2015, 2016, and 2017 in preparation of future operational forms. The RFP did not specify if the rangefinding and field test scoring in 2018 was part of the proposed contract scope of work. Given that the work is in support of future operational assessments and continues well past the end of the 2017-2018 fiscal year (year 4) our Cost Proposal assumes that this work will be done as part of a new contract or contract extension for year 5.

We will train and qualify PARCC readers on a baseline item for ELA/L or a prototype item for mathematics. Readers for ELA/L will then complete bridge training for the item they will score, and again for each successive item they score. Readers for mathematics will score their prototype item, then may continue on to score one or two similar items, completing a bridge set before scoring each successive item. If mathematics readers move to a different item type later in the scoring window, they will requalify on the appropriate prototype item.

Shortly after contract award, we will conduct a scoring planning meeting with PARCC staff to further discuss program requirements and to plan critical events, including rangefinding. After this meeting, we will develop a detailed rangefinding plan, as part of the handscoring specifications, documenting our assumptions, roles and responsibilities, schedule, the number of items and tasks we will review, the configuration of rangefinding sets, committee participants and agendas, and logistics, and submit this plan to PARCC for review and further input.

The rangefinding plan will include the sections and steps listed in the figure below.

Rangefinding Plan Components	
Section	**Steps**
Introduction	▪ List PARCC and Pearson staffs ▪ Document expected outcomes of the rangefinding session
Rangefinding Preparation	▪ Purpose ▪ Participant roles and responsibilities ▪ Overview of pre-rangefinding process ▪ Items and tasks ▪ Description or chart of responses needed per prompt ▪ Set labeling information ▪ Logistics
Rangefinding	▪ Overview of rangefinding process and meeting structure ▪ Security ▪ Daily agenda
Post-Rangefinding/Prior to Training	▪ Overview of the training development process ▪ Description or charts of responses/sets needed to satisfy training requirements ▪ Training Set Approval
Summary	▪ Review of the rangefinding plan ▪ Review of participants ▪ Expected outcomes

Rangefinding Plan Components. We will submit a rangefinding plan for ongoing field testing of new items to PARCC for review.

Security During Rangefinding

We will maintain stringent data security protocols throughout the preparation of the rangefinding materials, as well as during the meetings themselves. This includes training our employees to carefully follow security procedures and report security issues to the appropriate personnel.

Our documented security measures include the following:

- Storing all rangefinding materials in secure places and locking or otherwise securing meeting rooms overnight and during breaks
- All rangefinding materials will be numbered and accounted for at the conclusion of each session
- Archiving or shredding all excess photocopies and notes from each session
- Deleting or archiving unneeded electronic copies

The following sections will provide further detail on the rangefinding components of Pearson's scoring solution for PARCC Operational Assessments.

Conduct Field-Test Rangefinding Meetings

To facilitate productive field-test rangefinding meetings with PARCC educators and Partnership Management staff, Pearson will plan and organize meetings to select training papers. These meetings will:

- Take place in a hub city of a PARCC state, to be determined upon contract
- Involve up to 10 PARCC educators per grade level team, Partnership Management staff, and Pearson staff
- Be held concurrently when at all possible, over a period of up to three weeks
- Separate meetings four to five days in length will be conducted for each subject/grade combination
- Be conducted in two waves, one in the spring and the second in the fall
- Be chaired by either a Partnership Manager lead or a PARCC state lead

For each ELA/Literacy performance task, Pearson will bring samples of at least 100 student responses in order to provide a representative sample of responses for each task. For each Mathematics performance task, the number of sample student responses will be at least 120. Using the item/task rubrics, scoring directors will sort student responses by level of achievement. Pearson staff will group the responses into packages of twenty that represent the range of possible approaches and achievement levels (high, middle, low). Pearson will prepare 5-6 packets for each item, and duplicate those packets for each person participating in the rangefinding selection meeting.

Each student response copied for review by the committees will be assigned a unique number for the purposes of the rangefinding committee, and a corresponding log will be used to record important comments and decisions. The rangefinding committee will systematically review the photocopied responses to determine and record consensus scores and make recommendations for the possible placement of papers within training sets. Pearson will take notes during these meetings and document the rationale for score assignment.

By systematically reviewing student responses from each item, the rangefinding committees will set the standards for how the PARCC rubrics should be applied. Careful review of rubrics and student responses during rangefinding committee meetings and the subsequent compilation of anchor, training, and qualifying sets provides the initial step toward effective training and consistency of scoring.

Pearson staff will use the results from the rangefinding committee meetings to select training, qualifying, and validity responses, and will annotate training responses. Each anchor and training response will be annotated.

All selections, annotations, and other training materials will be approved by the Partnership Manager before reader training occurs.

Rangefinding Committee

Rangefinding meetings will be conducted by Pearson staff, and will also include Partnership Management staff and PARCC educators. The rangefinding meeting configuration is outlined below, and is shown in Table V.B.2.K.4.b of the RFP.

Rangefinding Committee Configuration	
ELA/Literacy	**Mathematics**
Grade 3	Grade 3
Grade 4	Grade 4
Grade 5	Grade 5
Grade 6	Grade 6
Grade 7	Grade 7
Grade 8	Grade 8
Grade 9	Algebra 1 / Mathematics 1
Grade 10	Geometry / Mathematics 2
Grade 11	Algebra 2 / Mathematics 3
Total membership = 90 ▪ 10 members per grade level team ▪ In each grade level team, there will be at least one lead	Total membership = 90 ▪ 10 members per grade/course ▪ In each grade/course team, there will be at least one lead

Rangefinding Meeting Requirements

Rangefinding meetings, as planned, organized, and conducted by Pearson, will meet the following stated requirements:

a. Pearson will comply with the requirements for Test Development Review Meetings as defined in Section V.A.1.G.3.1 on page 69 of the RFP.

b. We will include the appropriate Pearson staff for rangefinding meetings Pearson representation at rangefinding meetings will include handscoring leads who will facilitate discussions with PARCC representatives, take careful notes, and record consensus scores. In addition, Pearson will send content specialists (subject area leads) and program management support to every rangefinding event.

c. Pearson will arrange and pay for the meetings space, and manage the logistics for the committee meetings, including providing any media (for instance, computers, projectors to display rubrics, or other materials). We will arrange and pay for all rangefinding meeting travel for participants, including car rental and air transport, guest rooms, and reimbursement of travel expenses. We will provide breakfast and lunch during rangefinding meetings, as well as a partial per diem for dinner expenses.

d. We will pay travel costs for up to six PARCC Partnership representatives to attend each in-person rangefinding meeting.

Conduct Rangefinder Review Meetings

ELA/Literacy and Mathematics—Rangefinder Review Meetings

Three months prior to the beginning of handscoring training for the spring Operational Assessment test administrations, Pearson scoring staff will conduct a meeting in Washington DC to review rangefinder papers and scoring guides for the ELA/literacy and mathematics operational tests. The Pearson Project Lead and the subject/grade handscoring leads will collaborate with PARCC staff to review scoring criteria and rangefinder papers to score the tasks used during field test scoring or during the previous operational use.

Separate meetings will be conducted for each subject/grade combination. We will prepare necessary materials for the meetings.

These meetings will be used to review the rubric criteria, supplement the initial set of rangefinder papers with additional ones, if necessary, and to facilitate communication between PARCC Partnership Manager staff and Pearson scoring staff so that we share the same detailed understanding of the scoring criteria for the operational performance tasks. The scoring standards established by the initial field test rangefinder selection will be maintained during the subsequent rangefinder review.

Rangefinder Review Committee

The rangefinder review meetings will be attended by a rangefinder review committee, staffed by the Partnership. Pearson will anticipate the membership and configuration for this rangefinder review committee as outlined in the RFP and the figure below.

Rangefinder Review Committee Configuration	
ELA/Literacy	**Mathematics**
Grade band 3-5	Grade band 3-5
Grade band 6-8	Grade band 6-8
Grade band 9-11	Grade band 9-11
Total membership = 18 • 6 members per grade band • In each grade band, there will be at least one lead • In each grade band, at least one member will be an accessibility expert	Total membership = 18 • 6 members per grade/course • In each grade band, there will be at least one lead • In each grade band, at least one member will be an accessibility expert

Rangefinder Review Meeting Requirements

Rangefinder review meetings, as planned, organized, and conducted by Pearson, will meet the following stated requirements:

a. Schedule for rangefinder review meetings can be found in the meeting schedule. Pearson will comply with all requirements for test development review meetings as defined in Section V.A.1.G.3.1 on page 69 of the RFP.

b. We will include the appropriate Pearson staff for rangefinder review meetings. Pearson representation at rangefinder review meetings will include Handscoring leads who will facilitate discussions with PARCC representatives, take careful notes, and record consensus scores. In addition, Pearson will send content specialists (subject area leads) and program management support to every Rangefinding event.

c. Pearson will arrange and pay for the meetings space, and manage the logistics for the committee meetings, including providing any media (for instance, computers, projectors to display rubrics, or other materials). We will arrange and pay for all rangefinder review Meeting travel for participants, including car rental and air transport, guest rooms, and reimbursement of travel expenses. We will provide breakfast and lunch during rangefinder review meetings, as well as a partial per diem for dinner expenses.

d. We will pay travel costs for up to six PARCC Partnership representatives to attend each in-person rangefinder review Meeting.

Rangefinding/Rangefinder Review Participant Travel Costs

Pearson will be responsible for travel arrangements and travel costs associated with rangefinding and rangefinder review meetings for attending participants, including up to six Partnership representatives, as described in the meeting requirements and the schedule articulated below.

Rangefinding/Rangefinder Review Schedule

PARCC envisions rangefinding to support both operational scoring and field test scoring, with the latter logically divided into two phases. We have planned for rangefinding to take place according the following high level plan for ELA/literacy PCRs and mathematics items, with specific dates to be finalized in concert with the Partnership Manager and captured in the rangefinding plan (part of the handscoring specifications).

Rangefinding/Rangefinder Review Schedule					
Type	**Content**	**Timing**	**Anticipated Location**	**Partnership Manager Staff**	**Anticipated State Participants**
Phase 1 FT Rangefinding	ELA/L PCRs	May	Hub City	6	90
Phase 1 FT Rangefinding	Mathematics	May	Hub City	6	90
Phase 2 FT Rangefinding	ELA/L PCRs	Sept.	Hub City	6	90
Phase 2 FT Rangefinding	Mathematics	Sept.	Hub City	6	90
Operational Rangefinder Review	ELA/L PCRs	Nov.	Washington, DC	6	18
Operational Rangefinder Review	Mathematics	Nov.	Washington, DC	6	18

* Per Q & A 63, references to field test rangefinder meetings have been removed from Table.

** For additional details, see the Travel and Meeting tab in the Cost Proposal.

Produce Handscoring Materials

For each Operational Assessment performance task, Pearson will produce the following scoring materials, all of which will be approved by Partnership Manager staff prior to use:

- Scoring guides, containing item/tasks, passages (if applicable), rubrics; scoring criteria, glossary of key terms, and other material developed by Pearson and approved by the Partnership Manager. The scoring guide will include rangefinder materials, which will be approved by the Partnership Manager. Details on rangefinding materials may be found in the Rangefinding and Rangefinder Review section.
- Training sets, including annotated responses to be approved by the Partnership Manager and used for training practice. Training sets will be delivered online through the scoring system and will also be available to readers in their reference library throughout the scoring process. Details on training sets and papers may be found in the Recruit and Hire Readers section.
- Qualifying sets, presented to readers through the scoring system and later available to them in their reference library. Details on qualifying sets may be found in the Recruit and Hire Readers section.
- Validity sets, delivered to readers through the scoring system, including transcribed papers (if field test responses vary in mode to operational scoring), which will be proofread by qualified staff. Details on validity sets, validity response introduction processes, and designated staff responsibilities may be found in the Training and Qualifying Readers and Monitor and Maintain Handscoring Quality sections.
- Calibration sets, to provide additional training to all scorers at the same time. These will be delivered through the scoring system and later become part of the readers' reference library. Details on calibration sets and procedures may be found in the Monitor and Maintain Handscoring Quality section.

Pearson will catalog and store the current PARCC scoring materials upon delivery at the beginning of the project, along with all scoring materials throughout the course of the project. Pearson's content management system will be a valuable aid to managing the large amount of content to be developed, used, and stored throughout the life of this program.

Further details on the training materials Pearson staff will develop can be found in the Training and Qualifying Readers section below. We will submit all materials to the Partnership Manager for review and approval. The materials must be approved at least three weeks prior to the start of reader training and scoring. The schedule for these activities will be outlined in the overall project schedule and within the handscoring specifications. The figures below provide a summary of the handscoring materials for the Operational Administrations and field tests.

Operational Administration Prototype Training Sets								
	ELA Prose CR	Math - Type II & Type III	Algebra I - Type II & Type III	Geometry - Type II & Type III	Algebra II - Type II & Type III	Int Math I - Type II & Type III	Int Math II - Type II & Type III	Int Math III - Type II & Type III
Grades	3-11	3-8	HS	HS	HS	HS	HS	HS
Unique Item Count per Grade	3	18	21-25	21-25	27-30	3	3	3-4
Scoring Guides per Item	3	1	1	1	1	1	1	1
Anchor Papers per Set*	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt
Training Sets per Item	3	2	3	3	3	3	3	3
Training Papers per Set*	10	10	10	10	10	10	10	10
Qualification Sets per Item	3	3	3	3	3	3	3	3
Qualification Papers per Set	10	10	10	10	10	10	10	10
Rangefinder Review Days	15	15	15	15	15	N/A	N/A	N/A

Operational Administration Abbreviated Training Sets								
	ELA Prose CR	Math - Type II & Type III	Algebra I – Type II & Type III	Geometry – Type II & Type III	Algebra II – Type II & Type III	Int Math I – Type II & Type III	Int Math II – Type II & Type III	Int Math III – Type II & Type III
Grades	3-11	3-8	HS	HS	HS	HS	HS	HS
Unique Item Count per Grade	12	18	21-25	21-25	27-30	3-4	3-4	3-4
Scoring Guides per Item	1	1	1	1	1	1	1	1
Anchor Papers per Set*	16	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt
Training Sets per Item	2	2	2	2	2	2	2	2
Training Papers per Set*	10	10	10	10	10	10	10	10
Qualification Sets per Item	0	0	0	0	0	0	0	0
Qualification Papers per Set	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Rangefinder Review Days	15	15	15	15	15	n/a	n/a	n/a

FT Administration Training Sets								
	ELA Prose CR	**Math - Type II & Type III**	**Algebra I - Type II & Type III**	**Geometry - Type II & Type III**	**Algebra II - Type II & Type III**	**Int Math I - Type II & Type III**	**Int Math II - Type II & Type III**	**Int Math III - Type II & Type III**
Grades	3-11	3-8	HS	HS	HS	HS	HS	HS
Unique Item Count per Grade	42-48	52-78	110-162	110-162	114-162	27-36	27-36	28-36
Scoring Guides per Item	1	1	1	1	1	1	1	1
Anchor Papers per Set*	10	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt	3 per score pt
Training Sets per Item	1	1	1	1	1	1	1	1
Training Papers per Set*	5	5	5	5	5	5	5	5
Qualification Sets per Item	0	0	0	0	0	0	0	0
Qualification Papers per Set	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Rangefinding Days	15	15				n/a	n/a	n/a

Produce Handscoring Materials. Pearson staff will produce necessary scoring materials for each field test performance task. Rangefinding for the Integrated Mathematics will be included in the algebra I, algebra II, and geometry rangefinding meetings.

Delivery of Handscoring Materials

At the completion of scoring, we will provide PARCC with copies of all scoring materials prepared for and used during scoring. We will ship these materials to PARCC on labeled CDs stored within labeled cases by subject (and grade or mode as necessary). The CDs will include scoring summaries and calibration papers.

As noted in the Training and Qualifying Readers section and the Monitor and Maintain Handscoring Quality section, our robust scoring system allows for consistent, real-time access to scoring materials, including the Handscoring materials in the table above. Materials development, review, and approval can all be facilitated through the Content Management

System, which features an easy-to-use user interface for training set review and approval. The scoring system can deliver these final materials online, reducing waste and turnaround.

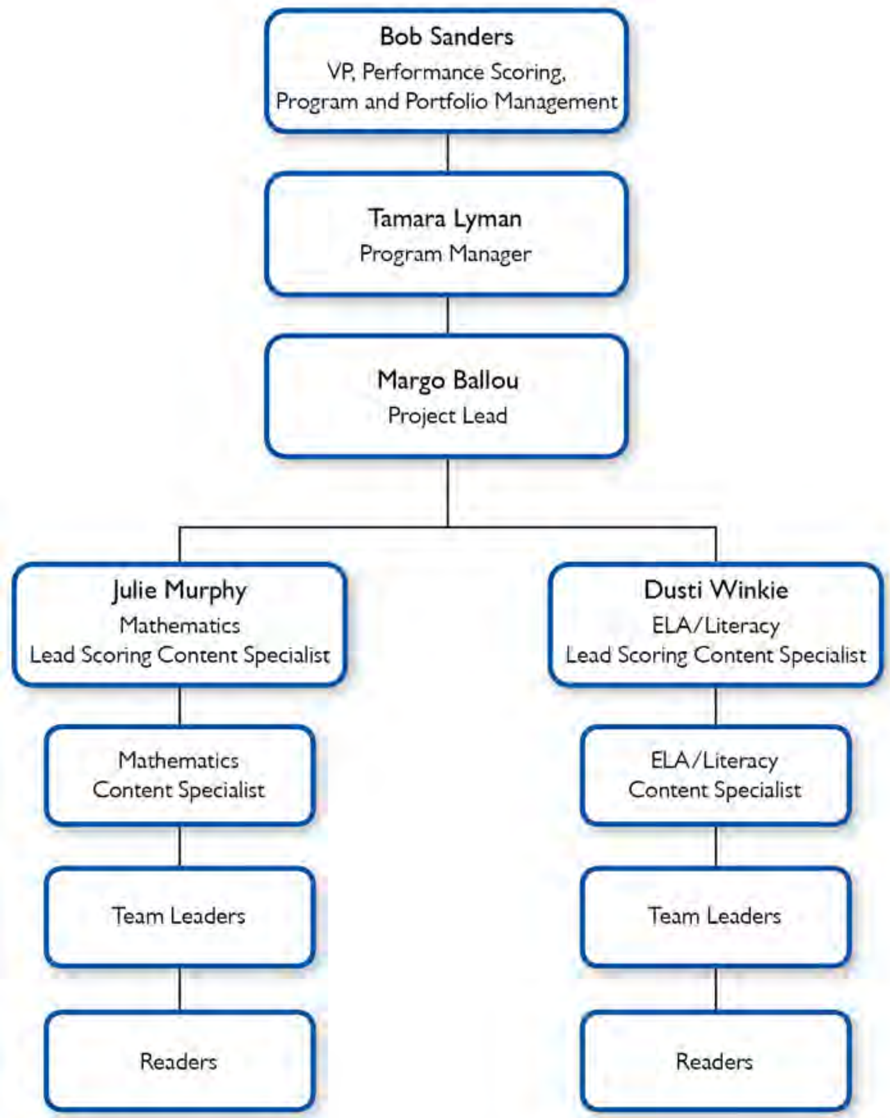
Contractor Staffing for Handscoring

Staffing the PARCC Program

Due to the size, complex nature, and aggressive schedule of the PARCC Operational Assessments, experienced key scoring staff will be vital to the program's delivery. The following figure shows key scoring staff who have been selected for their expertise and experience relevant to this critical, high-profile program.

Because our scoring staff works closely with Pearson's item development team, our staff members have reviewed and are familiar with newly emerging Common Core-aligned items, tasks, and rubrics for our state and national customers who are transitioning to the new assessments. Key staff members will be dedicated on a full-time basis to the management and implementation of the PARCC program.

Scoring leadership will be augmented by highly qualified professional readers, from our database of more than 100,000 screened applicants. An outline of our recruiting and staffing processes, from in-house scoring leaders through readers, is detailed in the Recruit and Hire Readers section below.



Scoring Support for the Operational Assessment. Pearson has the in-place staff, scoring knowledge, and expertise to successfully score the Operational Assessment.

Vice President, Performance Scoring, Program and Portfolio Management

Bob Sanders, Vice President of Performance Scoring, Program and Portfolio Management, plans and leads US-based performance scoring programs. Mr. Sanders provides strategic leadership to content, program, and scoring staff, and is accountable for scoring delivery. He has in-depth experience in many states, including overseeing staff managing handscoring programs in Arizona, Colorado, Washington, D.C., Florida, Illinois, Kentucky, Maryland, New York, and Ohio.

Mrs. Sanders is a certified Project Management Professional (PMP) by the Project Management Institute (PMI®). He has 15 years of experience in performance assessments, including ten years in performance scoring.

Performance Scoring Program Manager

Tamara Lyman, Scoring Program Manager for the Performance Scoring Center, will be responsible for planning, implementing, and overseeing professional human scoring required for the PARCC Operational Assessment. Ms. Lyman is the current scoring program manager for the PARCC Assessment Administration contract, working closely with PARCC, Inc., PARCC representatives, our partners, and our internal team.

Ms. Lyman was hired specifically to lead handscoring planning and implementation of the PARCC program, and will be fully dedicated to this program. Her existing experience with PARCC assessments includes the PARCC Rubric Study of the Item Development Research Studies contract; and the Item Review, Item Tryout, and current planning for the field test under the PARCC Assessment Administration contract. Her leadership, innovation, and creativity will be beneficial, as she will be responsible for coordinating the work of scoring project managers, content specialists, and handscoring leads assigned to the project. Ms. Lyman has over 25 years of experience in program and project management.

Performance Scoring Project Lead

Margo Ballou, Performance Scoring Project Lead for the Performance Scoring Center, will plan and implement scoring activities, including monitoring quality and schedule. The current Project Manager of field test scoring for the PARCC Assessment Administration, Ms. Ballou has planned and directed scoring at Pearson as Project Lead/Content Specialist, Scoring Director, and Assistant Scoring Director in the subjects of reading, writing, and adult literacy and numeracy. Her Pearson experience entails several large–scale assessment projects, including SAT, NAEP, and PARCC in the United States, and international projects such as the Programme for the International Assessment of Adult Competencies (PIAAC) and Progress in International Reading Literacy Study (PIRLS).

In addition to her current role on the field test of the Assessment Administration contract, Ms. Ballou has also served as project manager for the Rubric Study of the Item Development Research Studies contract; and the Item Review and Item Tryout of the Assessment Administration contract. Ms. Ballou has 11years of teaching experience at the secondary and postsecondary level, and 11 years of experience with handscoring programs.

Content Specialists

Our content specialists for the PARCC Operational Assessment have experience with the PARCC program, and will be directing the 2014 field test scoring of the PARCC Assessment Administration contract. Content specialists are content and scoring experts assigned as lead content staff to specific subject areas of a scoring project.

Both lead staff members assigned to the PARCC program are highly experienced senior scoring content specialists:

- Julie Murphy, senior scoring content specialist for Mathematics. Ms. Murphy supervises team leads, coordinates staff assignments and balances workloads, reviews all reports and evaluations regarding the performance of scoring staff, reviews evaluations of potential team leads, and supervises training and development of supervisors and readers for future content roles. Ms. Murphy has experience as a Senior Content Specialist, Content Specialist, Content Supervisor, Project Manager, Scoring Director, Scoring Supervisor, and Reader in mathematics and alternate assessments. Ms. Murphy currently serves as the Senior Content Specialist on the Item Review, Item Tryout, and field test of the PARCC Assessment Administration contract. Her 20 years of teaching and tutoring experience includes middle school and high school mathematics. She has nearly 14 years of experience in handscoring and assessments.
- Dusti Winkie, senior scoring content specialist for ELA/Literacy with twenty years of experience in the scoring industry. Ms. Winkie mentors team leads and monitors accuracy and consistency of content before and during rangefinding meetings, as well as during training development. She monitors quality and consistency during scoring across items and grades. Prior to joining Pearson, Ms. Winkie taught writing and German at the secondary and post-secondary levels and reading at the secondary level. She has served as Content Specialist, Scoring Director, Project Manager, and Scoring Supervisor on assessments in a variety of subject areas, with an emphasis on ELA. She spent five years working on new projects and has assisted with multiple research studies, including studies of the specific delivery models proposed for this program. Ms. Winkie currently serves as the Senior Content Specialist for the Item Review, Item Tryout, and field test of the PARCC Assessment Administration contract. Ms. Winkie has 20 years of experience in handscoring programs and 6 years of teaching experience.

Ms. Murphy will oversee mathematics scoring, and Ms. Winkie will oversee ELA/literacy scoring. Together, they will:

- Work with the program manager and project lead to complete sections of the hand-scoring specifications and other planning documents, including scoring design, range-finding and rangefinder review meetings, and training materials
- Attend range-finding to help monitor and maintain consistency
- Oversee handscoring leads and team leaders as they develop training based on the outcomes of rangefinding
- Guide the work of the team leaders across all phases of the scoring project
- Assist in maintaining accuracy and consistency across items and tasks
- Work with PARCC and Pearson scoring leadership staff and team leaders to monitor and maintain quality and consistency of scoring.

Ms. Murphy and Ms. Winkie will serve as lead content specialists, and will be supported by additional content experts in mathematics and ELA/literacy.

In some cases, including situations beyond Pearson's control, a named staff member may be unavailable at the time this program is implemented. In such cases, Pearson will provide another staff member with comparable experience or qualifications to fill the position described.

Team Leaders

Our content specialists will direct a team of scoring leaders comprising handscoring leads and team leaders.

Handscoring Leads

Two handscoring leads will be assigned for each grade and subject combination. Our handscoring leads, all of whom hold a Bachelor's degree or higher, have extensive experience overseeing large-scale scoring projects. They demonstrate proficiency in problem solving, decision making, training, and leadership skills.

Pearson handscoring leads have experience working with educators, whether in a rangefinding setting as meeting facilitators and record keepers, or in teacher workshop settings, including training and scoring sessions designed for educators. They rely on excellent customer service, public speaking, and organizational skills to yield positive, productive, and successful meetings with educators. For the Operational Assessment and ongoing field testing, handscoring leads will:

- Attend and facilitate rangefinding and rangefinder review meetings
- Develop and direct the training for team leaders and readers, based on knowledge gained at rangefinding
- Facilitate consistency with rangefinding decisions throughout training and scoring
- Maintain the consistency and quality of the scores assigned for the project
- Monitor scoring progress throughout the project

Team Leaders

Team leaders at Pearson may be either scoring directors with qualifications similar to those of our Handscoring leads, or they may be scoring supervisors. All scoring supervisors also hold a bachelor's degree or higher, have experience with scoring educational assessments, and must demonstrate strong communication, organizational, leadership, and decision-making skills. Team leaders work closely with handscoring leads to monitor reader training and scoring quality, answer questions, and evaluate their team members' performance. The Contractor will hire one team leader for every 10 to 12 active readers.

Employees selected to serve as team leaders for the Operational Assessment will:

- Demonstrate the ability to lead training and discussion sessions by successfully articulating scoring criteria and their proper application
- Demonstrate scoring expertise and the ability to monitor the accuracy of PARCC readers and respond with suitable feedback
- Assist handscoring leads and content specialists in reinforcing other project protocol and requirements

All team leaders will undergo the same training as readers, with additional team lead training, prior to their involvement in reader training. Training sessions will include training on scoring subject matter, as well as training on how to deal with readers' questions. Training for team leaders will include the successful completion of training sets and qualifying rounds for each subject.

Recruit and Hire Readers

Scoring quality starts with the recruitment process and extends through screening and placement (assigning readers to items based on their skills and experience), training, qualification, and scoring.

Pearson has a robust process for recruiting, screening, and onboarding highly qualified new and experienced readers. The depth of our scoring pool enables us to match reader education, background, experience, location, and preference to the specific requirements of each project. In 2013, our active readers had the following educational experience:

- 100 percent with bachelor's degree or higher
- 50 percent with master's degree or higher
- 7 percent with PhD
- 40 percent with teaching experience

Based on the test design proposed for this program, our staff assigned to PARCC have collaborated with our human resources personnel to build a comprehensive plan to score the PARCC assessments. The table below provides a detailed analysis of recruitment and hiring specifications for the PARCC 2014-2015 Operational Assessments.

Traditional Administration									
2015 PARCC Operational Assessment: English Language Arts / Literacy									
Grade	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Student Volume	1,075,207	1,085,253	1,080,827	1,071,557	1,073,061	1,060,066	525,305	496,192	454,832
Items per Student	3	3	3	3	3	3	3	3	3
2nd Score %	10%	10%	10%	10%	10%	10%	10%	10%	10%
Responses Scored	3,548,183	3,581,335	3,566,729	3,536,138	3,541,101	3,498,218	1,733,507	1,637,434	1,500,946
Scorers Recruited	2263	2286	2278	2580	2584	2551	1399	1322	1213
Scorers Hired	1886	1905	1898	2150	2153	2126	1166	1102	1011
Scorers Qualified (70%)	1,320	1,333	1,328	1,504	1,507	1,488	816	771	707
Scorers Qualified Less Attrition (90%)	1,191	1,202	1,197	1,357	1,359	1,342	736	695	637
Scorers per Shift	298	301	299	339	340	336	184	174	159
Average Daily Scorer Production	84,481	85,270	84,922	84,194	84,312	83,291	41,274	38,987	35,737
Team Leads	60	60	60	66	67	66	41	39	36
Team Leads per Shift	30	30	30	33	34	33	21	20	18
Active Scorers per Team Lead	10	10	10	10	10	10	9	9	9

* There are two team lead shifts and four scorer shifts per day.
** The term “active scorers” refers to the number of scorers logged in and scoring.
*** 10% of online ELA/L responses will receive a 2nd AI assigned score. 10% of paper ELA/L responses will receive a 2nd human assigned score. Volumes are subject to change based on State participation.

Traditional Administration						
2015 PARCC Operational Assessment: Mathematics / End-of-Course						
Grade	**3**	**4**	**5**	**6**	**7**	**8**
Student Volume	1,075,206	1,085,252	1,080,828	1,071,556	1,073,060	1,060,066
Items per Student	7	7	7	7	7	7
2nd Score %	10%	10%	10%	10%	10%	10%
Responses Scored	8,279,086	8,356,440	8,322,376	8,250,981	8,262,562	8,162,508
Scorers Recruited	796	803	800	1168	1169	1154
Scorers Hired	663	669	667	973	974	962
Scorers Qualified (70%)	530	535	533	778	779	769
Scorers Qualified Less Attrition (90%)	490	495	493	720	721	712
Scorers per Shift	123	124	123	180	180	178
Average Daily Scorer Production	197,121	198,963	198,152	196,452	196,728	194,345
Team Leads	30	31	31	40	40	40
Team Leads per Shift	15	16	16	20	20	20
Active Scorers per Team Lead	8	8	8	9	9	9
High School / End-of-Course	**Algebra I**	**Algebra II**	**Geometry**	**Int Math I**	**Int Math II**	**Int Math III**
Student Volume	393,978	372,144	341,124	131,326	124,048	113,710
Items per Student	8	10	8	8	8	10
2nd Score %	10%	10%	10%	10%	10%	10%
Responses Scored	3,467,006	4,093,584	3,001,891	1,155,669	1,091,622	1,250,810
Scorers Recruited	548	842	736	184	246	281
Scorers Hired	457	702	613	153	205	234
Scorers Qualified (70%)	365	561	490	122	164	187
Scorers Qualified Less Attrition (90%)	338	519	453	113	151	173
Scorers per Shift	85	130	113	28	38	43
Average Daily Scorer Production	82,548	97,466	71,474	27,516	25,991	29,781
Team Leads	26	37	31	6	8	9
Team Leads per Shift	13	19	16	3	4	5
Active Scorers per Team Lead	7	7	7	9	10	9

* There are two team lead shifts and four scorer shifts per day.

** The term “active scorers” refers to the number of scorers logged in and scoring.
*** Volumes are subject to change based on State participation.

Spring Administration									
2015 PARCC Operational Assessment: English Language Arts / Literacy									
Grade	**9**	**10**	**11**	**HS**	**HS**	**HS**	**HS**	**HS**	**HS**
Content	**ELA/L**	**ELA/L**	**ELA/L**	**Algebra I**	**Algebra II**	**Geometry**	**Int Math I**	**Int Math II**	**Int Math III**
Student Volume	13,824	13,058	11,970	10,368	8,976	9,794	3,456	3,264	2,992
Items per Student	3	3	3	8	10	8	8	8	10
2nd Score %	10%	10%	10%	10%	10%	10%	10%	10%	10%
Responses Scored	45,619	43,091	39,501	91,238	98,736	86,187	30,413	28,723	32,912
Scorers Recruited	101	95	89	61	92	79	23	29	32
Scorers Hired	84	79	74	51	77	66	19	24	27
Scorers Qualified (70%)	63	59	55	42	63	54	15	19	22
Scorers Qualified Less Attrition (90%)	56	53	49	38	58	50	13	17	20
Scorers per Shift	14	13	12	10	15	13	3	4	5
Average Daily Scorer Production	3,259	3,078	2,822	9,124	9,874	8,619	3,041	2,872	3,291
Team Leads	4	4	4	6	7	6	5	5	6
Team Leads per Shift	2	2	2	3	4	3	3	3	3
Active Scorers per Team Lead	7	7	6	3	4	4	1	1	2

* There are two team lead shifts and four scorer shifts per day.
** The term “active scorers” refers to the number of scorers logged in and scoring.
*** 10% of online ELA/L responses will receive a 2nd AI assigned score. 10% of paper ELA/L responses will receive a 2nd human assigned score. Volumes are subject to change based on State participation.

Winter Administration									
2015 PARCC Operational Assessment: English Language Arts / Literacy									
Grade	**9**	**10**	**11**	**HS**	**HS**	**HS**	**HS**	**HS**	**HS**
Content	**ELA/L**	**ELA/L**	**ELA/L**	**Algebra I**	**Algebra II**	**Geometry**	**Int Math I**	**Int Math II**	**Int Math III**
Student Volume	19,824	19,028		13,824	13,058	11,968			
Items per Student	3	3		8	10	8			
2nd Score %	10%	10%		10%	10%	10%			
Responses Scored	65,419	62,792		121,651	143,638	105,318			
Scorers Recruited	149	144		70	107	122			
Scorers Hired	124	120		66	89	102			
Scorers Qualified (70%)	92	90		54	73	84			
Scorers Qualified Less Attrition (90%)	82	80		50	67	77			
Scorers per Shift	21	20		13	17	19			
Average Daily Scorer Production	4,673	4,485		12,165	14,364	10,532			
Team Leads	4	4		6	7	6			
Team Leads per Shift	2	2		3	4	3			
Active Scorers per Team Lead	11	10		4	4	6			

* There are two team lead shifts and four scorer shifts per day.
** The term “active scorers” refers to the number of scorers logged in and scoring.
*** 10% of online ELA/L responses will receive a 2nd AI assigned score. 10% of paper ELA/L responses will receive a 2nd human assigned score. Volumes are subject to change based on State participation.

Recruiting Process

Our in-house human resources staff understands the scoring business; we are not dependent on third-party temporary agencies that often lack the specific knowledge and sensitivity necessary to identify readers with the right qualifications. In 2015, our human resources staff plans to recruit nearly 32,000 readers to accommodate the PARCC Operational Assessment. With 100,000 readers in our screened applicant pool, we have an extensive basis to draw from. Our team is also planning an outreach campaign, starting in 2014, specifically for the PARCC program. Our staff members are seasoned at using a diverse array of media channels to conduct national campaigns.

In recruiting and selecting readers for the PARCC Operational Assessment, priority will be given to individuals with previous experience scoring the PARCC field test in 2014, experience in scoring assessments in PARCC member states, and experience scoring Common Core-aligned items and tasks.

All readers will have degrees in mathematics, reading, education, or a related field. Readers of mathematics performance task responses will have completed a four-year college degree program and have the mathematics knowledge needed to effectively score responses to mathematics items. Readers for ELA/literacy with have degrees in reading, education, history, psychology, journalism, or a related area and/or teacher certification that will enable them to succeed scoring the literacy analysis, research simulation, or narrative writing tasks. All degrees are verified by a third party, and potential readers whose degrees cannot be verified before the start of live scoring will be dismissed.

In most cases, our professional readers possess specialized educational and professional experience, including valuable experience in performance scoring.

Once selected and placed on an assignment, readers will complete all necessary new-hire paperwork and an online orientation from home prior to their first day of employment. Readers will be ready to begin training on and scoring PARCC responses upon starting the project. Applicants hired for training must also sign an agreement with the Partnership Manager that they will maintain the security of PARCC's test materials in addition to security agreements required by the Contractor.

The following figure provides an overview of the key process steps and timelines.

Recruiting and Selecting Team Leads and Readers	
Process	**Timeline**
Recruitment Tool A web-based application used to collect data on reader education, prior scoring experience, teaching credentials, work status, and other key information to screen candidates.	Throughout the year
Initial Screening Candidate data are analyzed and prospective readers prioritized.	Six to eight weeks prior to project start
Interviews Telephone or online interviews conducted to collect additional data for reader screening and placement.	Six to eight weeks prior to project start
Offer Offer letters emailed to prospective readers, detailing project requirements, timelines, and quality standards contingent upon proof of degree.	Upon successful screening and interview
Verification Degrees verified through the National Student Clearinghouse or through the institution itself. Prior experience is provided through hard copy documentation.	Six to eight weeks prior to project start

Final Documentation and Project Placement Readers sign confidentiality agreement in which they agree to keep all information and student responses confidential. Only readers who successfully complete onboarding, training, and qualifying are allowed to evaluate student responses.	Upon hire

Recruiting and Selecting Readers for the PARCC Operational Assessment. Our recruiting and selection process is based on our many years of experience in the scoring industry.

Qualified Readers. Pearson places emphasis on recruiting and hiring a large pool of degreed, qualified individuals to score assessments through distributed scoring and in regional scoring hubs. In order to provide PARCC with highly competent readers for the Operational Assessment, potential readers will demonstrate their qualification to score through successful completion of training and meeting the standard on qualification rounds, as indicated in the handscoring specifications document.

Qualification File. We will document qualifications of readers in a Qualification File. This file will indicate degrees earned; relevant teaching, educational, or work experience; and previous scoring experience relevant to scoring PARCC member states' assessments. This file will not include personally identifiable information, such as name, address, and telephone number. After scoring has begun, we will only maintain Qualification Files for readers who successfully qualify and score student responses to PARCC assessments.

In the month prior to handscoring training, we will provide for the PARCC Partnership Manager's review, weekly reports of the progress of potential reader recruitment efforts as well as the qualifications of potential Team Leaders.

All potential readers will sign a statement indicating that they understand the following conditions:

- If applicants do not successfully complete the training and qualifying requirements, they will not be hired as readers
- If they are hired as readers, they may be dismissed if, after being trained to score, their scoring performance does not meet the requirements of the Partnership Manager or Pearson
- While the distributed scoring model allows us to access a large pool of highly talented and qualified readers, many of these readers complete their scoring responsibilities in a part-time capacity, as many are full-time teachers.

Applicants will also sign an agreement with the Partnership Manager that they will maintain the security of PARCC's test materials in addition to security procedures required by Pearson.

We will provide independent, third-party verification that potential readers for the PARCC Program have earned their stated college degrees.

Handscoring Monitoring. As we are recommending a distributed scoring model for the majority of the Operational Assessment scoring, monitoring will take place continuously throughout the scoring process. Details on the system's monitoring capabilities can be found in the Handscoring Solution Overview section of this document.

For regional scoring hubs, Pearson will provide oversight and monitoring of team lead and reader qualification prior to and during training and qualification processes.

Pearson will be responsible for making car rental and air transport arrangements for Partnership Management representatives observing and monitoring handscoring, securing guest rooms for handscoring monitors near the handscoring site, and reimbursing the Partnership Management staff for all travel-related expenses. The Contractor will provide a meal per diem payment to each Partnership Management staff.

Monitoring of Regional Handscoring Hub Sites				
Meeting	**Meetings per Year**	**Days per Meeting**	**Number of Partnership Manager Staff**	**Number of State Participants**
Handscoring Monitoring	3	10	5	2

Monitor Hub Sites. Partnership Management will be able to observe training and scoring in the regional hub sites.

Training and Qualifying Readers

The first step toward accurate and consistent scoring of the Operational Assessment is accurate, consistent, and thorough reader training. Readers hired for scoring of the PARCC Operational Administration will go through rigorous, project-specific training. Pearson readers are coached and monitored throughout training, and they must qualify to score based upon the Partnership Manager's standards. Each project has its individual properties, including reader qualification, and those are adhered to throughout the project.

Pearson's expert staff, many of whom have an education or teaching background and familiarity with creating effective learning environments, has many years of experience designing and developing robust, effective training for readers and supervisors. Our content staff will work closely with Partnership Management staff to develop effective training for the Operational Assessments.

Pearson's scoring staff recognizes the need to be flexible in creating and revising training materials, based upon the availability of suitable responses for each training segment. We strive to build training plans that ground readers in the PARCC standards and guide them throughout the scoring process.

During training set building, our team leaders will focus on maintaining the decisions and intent of the committees and Partnership Manager staff. Final training sets are described in the following figure.

Operational Assessment ELA/Literacy Baseline Training Sets	
Specification	**Description**
Scoring Guide and Anchor Set	
Scoring guides will be produced for each reader to use during training and scoring. The scoring guides will consist of the rubrics, scoring criteria, performance tasks, passages (if applicable), glossary of key terms, and other scoring guidelines designated by the Partnership Manager and Contractor. The scoring guides will also contain anchor sets consisting of two to four rangefinder responses for each score point for mathematics, and a minimum of three responses for each score point for ELA/literacy. ELA/literacy will have three anchor sets, one for each trait. Each of the rangefinder responses in the anchor sets will be annotated. Contractor will be responsible for writing the annotations for the rangefinder papers.	The anchor set is the primary reference for readers as they internalize the rubric during training. All readers have access to the anchor set whenever they are training and scoring, and are directed to refer to it regularly. The anchor set comprises clear examples of student performance at each score point. The responses selected should be representative of typical approaches to the task, arranged to reflect a continuum of performance.
Practice Training Sets	
Three training sets of 10 responses will be developed for each ELA/Literacy performance task. For lower grade mathematics performance tasks, two training sets of 10 responses will be developed, and three training sets at the high school level. Responses in the training sets will be annotated for readers.	Practice sets are used to help trainees develop experience in independently applying the scoring guide or rubric to student responses. Some of these responses clearly reinforce the scoring guidelines presented in the anchor set. Other responses are selected because they are more difficult to evaluate, fall near the boundary between two score points, or represent unusual approaches to the task. The practice sets provide guidance and practice for trainees in defining the line between score points, as

Operational Assessment ELA/Literacy Baseline Training Sets	
Specification	**Description**
	well as applying the scoring criteria to a wider range of types of responses.
Qualifying Sets	
Three qualifying sets will be developed for each performance task. For mathematics and ELA/literacy performance tasks, qualifying sets consist of three sets of 10 responses for the baseline or prototype task for which the set is built.	Qualifying sets are used to confirm that reader trainees have grasped the scoring criteria and are able to assign the range of scores to student responses accurately. The responses in these sets are selected to clearly reinforce the application of the scoring criteria illustrated in the anchor set. Reader trainees must demonstrate acceptable performance on these sets by meeting a pre-determined standard for accuracy in order to qualify to score the field test.

Operational Assessment Abbreviated Sets	
Specification	**Description**
Scoring Guide and Anchor Set	
Scoring guides will be produced for each reader to use during training and scoring. The scoring guides will consist of the rubrics, scoring criteria, performance tasks, passages (if applicable), glossary of key terms, and other scoring guidelines designated by the Partnership Manager and Contractor. The scoring guides will also contain anchor sets consisting of two to four rangefinder responses for each score point for mathematics, and a total of 16 responses per anchor set for ELA/literacy. Each of the rangefinder responses in the anchor sets will be annotated. Contractor will be responsible for writing the annotations for the rangefinder papers.	The anchor set is the primary reference for readers as they internalize the rubric during training. All readers have access to the anchor set whenever they are training and scoring, and are directed to refer to it regularly. The anchor set comprises clear examples of student performance at each score point. The responses selected should be representative of typical approaches to the task, arranged to reflect a continuum of performance.
Practice Training Sets	

Operational Assessment Abbreviated Sets	
Specification	**Description**
Two training sets of 10 responses will be developed for each mathematics and ELA/Literacy performance task. Responses in the training sets will be annotated for readers.	Practice sets are used to help trainees develop experience in independently applying the scoring guide or rubric to student responses. Some of these responses clearly reinforce the scoring guidelines presented in the anchor set. Other responses are selected because they are more difficult to evaluate, fall near the boundary between two score points, or represent unusual approaches to the task. The practice sets provide guidance and practice for trainees in defining the line between score points, as well as applying the scoring criteria to a wider range of types of responses.
Qualifying Sets	
Abbreviated training materials will not include qualifying sets.	

Comprehensive Training. To provide quality scoring for the Operational Assessment, we will develop illustrative anchor, practice, and qualifying sets.

Online Training

For the PARCC Operational Assessments scoring, we propose online training of readers, which can offer advantages over traditional instructor-led training. We first implemented online training in 2002 with a Utah scoring project. Online training allows for more customer involvement in training development and for flexible yet consistent training that allows readers to train at their own pace.

Rigor and Flexibility of Online Training

The use of effective online training for the Operational Assessments will enable distributed scoring and the use of readers or teachers with specialized backgrounds when desired. We will deliver online training though our secure digital scoring system, providing an efficient and effective means of preparing readers to accurately and consistently score to PARCC standards.

Training includes modules that explain the prompt, rubric, scoring criteria, reader bias prevention, and all scoring decisions. It also includes fully annotated anchor sets and training sets, and qualifying sets that readers access at their own pace and from their own locations.

Given the high volume and complexity of the PARCC assessment program, the value of rigorous, consistent online training that will support high volumes of readers cannot be understated. The figure below summarizes key benefits of online training.

Benefits of Online Training to the PARCC Operational Assessment		
Feature	**Advantage**	**Benefit**
High-quality, customized training	Incorporates adult learning and instructional design principles in training and interactions. Item-specific training is customized to meet needs of PARCC states.	Readers receive training designed to engage, instruct, and provide frequent feedback on understanding, accuracy, and consistency in scoring the Operational Assessment
Enhanced Consistency	Helps maintain consistent standards across readers and locations.	All readers receive identical instruction, enhancing consistency between groups of readers.
Technology-enhanced Learning	Allows interactions to engage trainees, reinforce standards, and give trainees immediate feedback on understanding of concepts and standards.	Trainees receive clarification on concepts and scoring criteria as they train and before they reach the qualification phase, cementing correct understanding and leading to accurate and consistent scoring.
Individual Focus and Pace	Allows trainees to set their own pace for training, enhancing individual focus and comprehension.	Each trainee can focus on particular papers or features that are problematic, rather than having discussion determined by the needs of the group.
Efficiency	Automates training and monitoring processes that are otherwise manual and time-consuming.	Many of the clerical processes are automated.
Flexibility to Meet Schedules	Allows the ability to move items from one group of readers to another without jeopardizing the consistency of training or moving trainers from one site to another.	Readers train at their computer workstations, eliminating the need for separate training rooms, hotel meeting rooms, or convention centers.
Quickly Train Additional Readers	Allows readers trained after initial scoring has started to receive instruction identical to that of the original group of readers.	All readers receive identical instruction.

Online Training Benefits to the Operational Assessment. Potential PARCC readers will receive thorough training for the Operational Assessment.

Online Training Modules

Since our first distributed scoring project in 2002, we have continually refined and strengthened our online training. Each year, Pearson's content staff writes and our Course Design and Development (CDD) group develops and deploys more than 900 online training modules.

Our online training modules will explain, illustrate, and reinforce the standards and decisions established during PARCC rangefinding sessions. Readers encounter training activities that stress and reinforce key concepts and important scoring decisions. The figure below provides an example of the training module view.

Online Training Modules. Pearson's Content Staff and Course Design and Development group will create and implement effective and efficient training modules for PARCC readers.

Readers are encouraged to discuss questions or concerns with their team leader to receive one-on-one support, whether in a regional or distributed setting. Handscoring leads and team leaders mentor and monitor readers throughout the rigorous training, and are available to answer reader questions and clarify any points.

Because our content staff monitors readers closely throughout training and scoring, they are able to intervene, coach, and answer questions promptly. Training set scores can reveal patterns in misapplication of the rubric so that a reader's understanding of PARCC standards can be recalibrated before live scoring begins.

This attention to creating robust, accurate training and providing feedback during training will provide PARCC with the results they are seeking—namely, readers that understand and can apply the standards Partnership Manager staff have defined.

Our rigorous training protocol was developed by experienced handscoring leads, and provides consistency and predictability to all projects. In addition, the Partnership Manager will approve student responses, corresponding scores, and explanatory annotations contained in the training modules prior to use. As shown in the figure below, training is delivered in a modular format, which follows the reader's learning process, with interactivity to assess understanding and progress.

Online Training Modules	
Module	**Content**
Scoring for Pearson	Includes an introduction to Pearson; trains readers on appropriate Pearson policies and confidentiality requirements; and educates readers on the philosophy of scoring.
Pearson Scoring System	Readers learn how to submit scores and practice scoring in Pearson's scoring system.
Scoring the Project	Oftentimes presented as subject-level training, this module provides project-specific training, including an overview of the project, details on who will be supporting readers during the training and scoring process, and training on preventing reader bias they may encounter while scoring.
Scoring the Item (Scoring Guide)	Item-level training, this module provides readers with the specific requirements needed to accurately score the item they are assigned. Item-specific content includes level-setting the reader on the age of and resources available to the respondent, the prompt, rubric, scoring decisions, and annotated anchor papers.
Practice Scoring (Training Sets)	Provides readers with practical experience applying the scoring guidelines to sample student responses and includes feedback on their scoring.
Qualification (Qualification Sets)	Tests readers on the retention of the training and their ability to apply the rubric and standards to score accurately.
Pearson Scoring System Part Two	Provides readers with training on additional functionality of the scoring system, including how to communicate with supervisory staff and use of self-monitoring tools.
Before You Score	Presents advanced training on project-specific handling of responses that may require condition codes or alerts. This module also includes information to readers on monitoring and managing their scoring quality.

PARCC Assessment Training Modules. To achieve quality scoring for the Operational Assessment, Pearson readers are intensively trained.

Qualification of a reader is based upon a reader's ability to meet agreed upon standards. These standards will be published in the handscoring specifications. Qualified readers may score sets of items or single performance tasks. Pearson will work with the Partnership Manager to determine the most efficient method for scoring.

Team Leader Training

Prior to reader training, Pearson team leaders will complete online training modules, including training and qualifying sets, as well as training on working with readers and how to evaluate responses and scores. This training will provide team leaders with a clear understanding of the materials and scoring protocols.

Team leaders will carefully read and review annotations of all training materials. Handscoring leads will provide appropriate focus and emphasis so that the team leaders can assist in reader training and respond to readers' questions during training and scoring.

Team leaders receive system training, focusing on the use of our digital scoring system tools and applications. They receive instruction from handscoring leads on accessing, reading, and interpreting statistical reports generated by the system so they can help monitor trends and backreading needs.

Reader Training

Much like team leaders, readers will proceed through online training and qualifying prior to scoring. This training provides readers with a clear understanding of the training materials and scoring protocols of the PARCC Operational Assessment.

As with team leaders, readers will carefully read and review annotations of the training materials, with focused direction given by the handscoring leads and team leaders. We encourage readers to contact supervisory staff if they have questions during training. We also carefully monitor practice set reports during training and have the ability to contact readers through the scoring system to give instructions or feedback on trends.

Qualification

Reader trainees will be required to achieve the minimum perfect plus adjacent agreement percent on the qualification sets to qualify to begin scoring. If applicants do not successfully meet the training and qualifying requirements established by PARCC Partnership Manager staff and Pearson, they will not be allowed to score any Operational Assessment student responses. Furthermore, qualified readers may be dismissed if their scoring performance does not meet the standards set by PARCC. Details on training and qualifying sets for reader qualification can be found in the Handscoring Materials section of this response.

Reader Assignment

Following training and qualification on ELA/literacy and mathematics, readers will be assigned to performance tasks for the scoring of the Operational Assessment. These assignments will be based upon consultations between Pearson staff and Partnership Management staff.

Handscoring Reports

Our digital scoring system automatically captures and tracks all score data. By reviewing up-to-date reader performance statistics, we can quickly identify particular readers whose performance falls outside of group norms while also keeping close track of the group as a whole.

Reports for use in quality monitoring and project completion status are generated and updated automatically and are available to Pearson scoring leadership staff at any time via our digital scoring system. Our reports give daily and cumulative statistics by item and provide individual and group average agreement percentages. These reports will be shared with the Partnership Manager staff on a daily basis, and will be available for both operational and field-test scoring. The following figure summarizes the Handscoring Reports we will use for monitoring PARCC Operational Assessment scoring.

Reader Performance Reports		
Report Content	**Report Name**	**Description**
Reader performance on training sets	Cumulative Practice Report by Reader	Shows the following data from Practice Scoring from the start of the Project to the time the report is run: ▪ Per reader, his or her agreement percentages with corresponding True Scores for a given trait, for a given set ▪ Per reader, his or her agreement percentages with corresponding True Scores for a given trait, across sets ▪ Per item, all readers' agreement percentages with corresponding True Scores for a given trait, across sets
Reader performance on qualifying sets	Cumulative Qualification Report by Reader	▪ Shows the following data from Qualification Scoring from the start of the Project to the time the report is run: ▪ Per reader, his or her scores and their agreement percentages with corresponding True Scores for a given set ▪ Per reader, his or her agreement percentages with corresponding True Scores for a given trait, across sets ▪ Per item, all readers' agreement percentages with corresponding True Scores for a given trait
Reader performance on student responses	Reader Performance Report	Shows the scoring activity of all readers whose performance does not meet a selected threshold for Inter-Rater Reliability Agreement Percentage and/or Validity Agreement Percentage.
Reader performance on validity papers	Daily/Cumulative Validity Summary	Shows on both a daily and a cumulative basis the room-level summary of agreement for validity reads of a given item.
Reader performance on IRR	Daily/Cumulative Inter-Rater Reliability Summary	Shows the group-level summary of both daily and cumulative inter-rater reliability statistics for each day of the project.
Reader daily score point distribution	Daily Inter-Rater Reliability and Frequency Distribution by Reader	Shows the inter-rater agreement percentages and score point distributions for each reader for a given day.
Reader performance on read-behinds	Cumulative Backreading for Readers	Shows, for each reader: ▪ The number of backreads completed ▪ The overall percentage of agreement between readers' assigned scores and the Supervisor's backreading scores ▪ The percentage of adjacent and non-adjacent (low and high) assigned scores as compared to Supervisors' backread scores
Reader time spent scoring tasks or task sets	Team Leader Performance Peek	Snapshot of reader performance and time logged in scoring and/or training

Reader Performance Reports. A range of statistical reports on reader performance allows Pearson and the PARCC Partnership Manager to monitor real-time scoring trends throughout the project.

Handscoring Reports will include summary information by task, team, and site, as well as a summary of daily and cumulative numbers of student responses read and remaining reads. Our reports will allow for analysis of project progress and timeliness. Reports shall aggregate data by item (there will be no state-specific break-outs of data).

We will provide the Partnership Manager with an annotated list of all available reports at the onset of the first handscoring planning meeting, and will work with the Partnership Manager to determine which reports will be most useful and relevant for use during the Operational Assessment scoring. We will provide the Partnership Manager will final copies of all cumulative Handscoring reports at the close of operational and field test scoring in the file format requested or on a CD ROM as indicated in the RFP.

Details of our standard handscoring reports can be found in the Reader Performance Reports table in the Monitor and Maintain Handscoring Quality section below.

Scoring Student Responses

The PARCC assessment program requires a scoring approach that balances cost effectiveness with reliable, high quality results. To meet the needs of such an expansive and unique program, we recommend a distributed scoring model, supplemented by regional scoring hubs in participating PARCC states.

Operational Scoring

Scoring constructed response items for the Operational Assessment will be carried out through a distributed scoring model. Student responses will be read and scored independently by trained and qualified readers or the AI scoring engine, as illustrated in the table in Scoring Rules for PARCC Assessments section above. All student responses will receive a first score. Ten percent of the student responses will be automatically and randomly distributed for independent second readings. If the two readers' scores are nonadjacent, a third and sometimes a fourth reader shall be used.

Readers will have no indication of whether a response was scored previously, and they will not have access to any previous scores. The scoring system automatically routes responses requiring resolution reads to qualified readers.

Pearson acknowledges that some states may require 100 percent double scoring for high stakes tests for students (e.g., high school exit tests). As directed, we have provided costs for this alternate double scoring within the cost proposal in Section V.G. Our scoring system easily accommodates any variation of double scoring rules, which can be set up for each state's needs.

Field Test Scoring

Scoring of the PARCC field test items will be very similar to the operational administrations, as readers from the operational scoring will be used to score the field test items. We will use the same scoring system and quality monitoring tools for the field test as used for the Operational Assessments.

For the Field Test scoring, the percentage of second scoring for online ELA/L responses will be amended to accommodate the number of second scored responses needed to train the AI scoring engine (please refer to the footnotes in the revised Field Test Scoring Specifications figure).

Readers receive randomized individual student responses across randomized schools. The second read is also randomized and will not be scored by the initial reader. Pearson will work with the Partnership Manager to document all the scoring rules and these will be in the handscoring specifications which are reviewed and approved by the Partnership Manager.

Discussion Sets

The calibration tool discussed below in the Monitor and Maintain Handscoring Quality section will serve the purpose of discussion sets.

Alert Systems

Hand-in-hand with delivering accurate scores to PARCC, Pearson is also concerned with the safety and well-being of all students in PARCC member states. We train our readers to be aware of student responses indicating potential need for intervention.

Alerts include student responses that might indicate that a child is a danger to him or herself or others, is experiencing depression, is involved in abuse, or is contemplating suicide. Other alerts include responses that might indicate a testing irregularity may have occurred. We will follow our standard process for alerting such responses.

Scoring Rules

As outlined in PARCC Readers and Scoring Rules subsections of the Handscoring Solution Overview section above, student responses for the PARCC Operational Assessment will be scored independently by trained and qualified readers or by the automated scoring engine. Our digital scoring system automatically and randomly distributes student responses to readers (or the automated scoring engine) based on reader qualification and project requirements. Readers must confirm the entered score as valid prior to submission. Readers will have no indication of whether a response was scored previously, and they will not have access to any previous scores.

System Display

For both paper-based and computer-administered tests, the Pearson scoring system will present questions and responses to the reader, along with scoring support materials. For scanned tests, this delivery is based on the clip area designation.

During the scanning process, we create individual records for each student answer document containing the data captured during scanning, including PARCC state-provided student identification information and responses to multiple-choice and constructed-response items. We capture a full-image scan of a student document, create a clipped version of it, and send it to our performance scoring system.

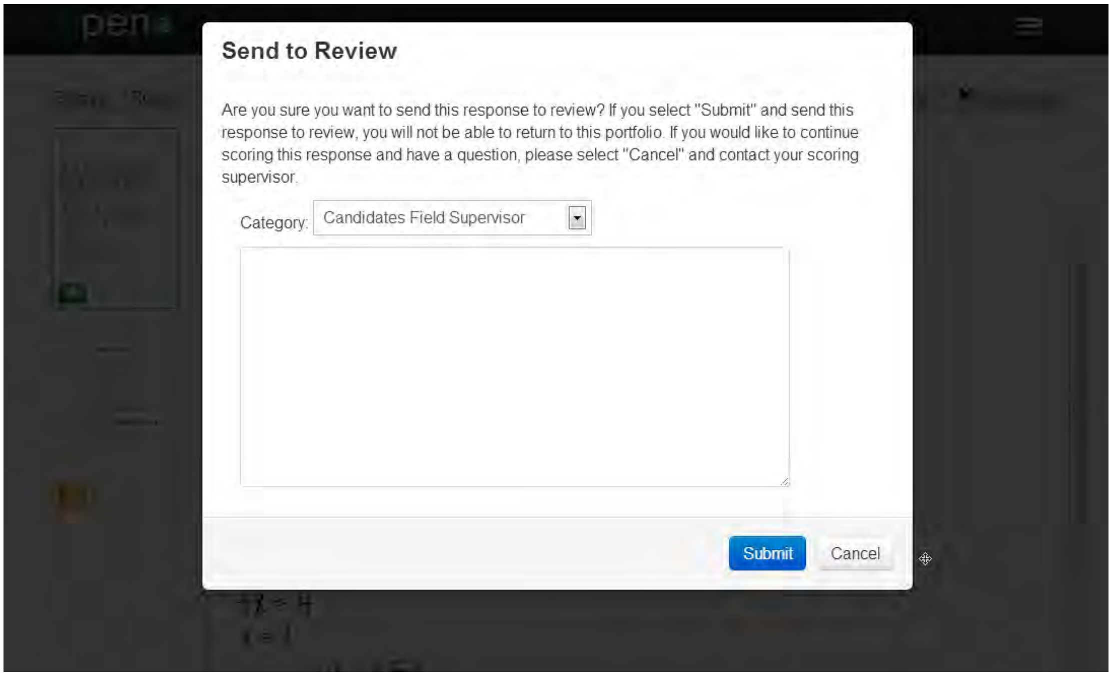
In the case of computer-administered tests, student responses are delivered as completed by the student, along with supporting materials, through the system to the reader.

Screen Display Features

Several features are available to readers using the Pearson system. The area of the screen where the reader applies scores is defined per item during setup. During the setup definition phase, PARCC and Pearson will work together to document and agree upon the options for the reader's view of the screen.

The figures below show several scoring features, including:

- **Scrolling Image** allows a reader to scroll a student's response when it is larger than the screen size.
- **Zooming Response Image** permits readers to increase or decrease response size.
- **Prompt and Rubric** are available for each response scored.
- **Refer to Anchor Set** is the primary reference for readers as they internalize the rubric during training. All readers have access to the anchor set whenever they are training and scoring, and they are directed to refer to it regularly. The anchor set includes clear examples of student performance at each score point.
- **Request a Review** enables readers to send a response to a supervisor for review, as needed. After a reader enters and submits the score, the response will be sent on for either a final score or additional routing, as determined by PARCC rules. Throughout the scoring process, reports confirm that responses are moving through the system according to the defined scoring rules.



Request a Review. Readers who have concerns or questions on a student response will be able to quickly submit the response for review within the scoring system.

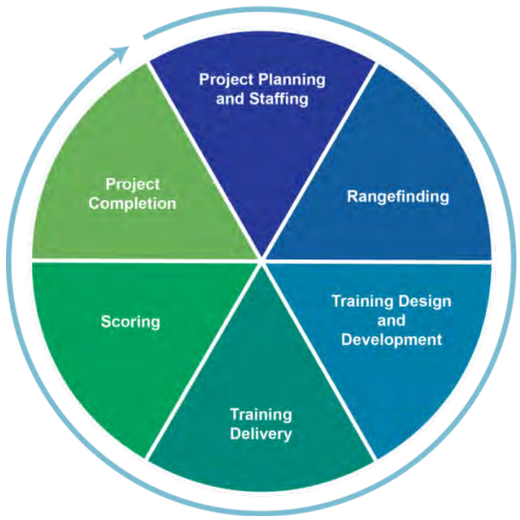
After a reader enters and submits the score, the response will be sent on for either a final score or additional routing, as determined by PARCC rules. Throughout the scoring process, reports confirm that responses are moving through the system according to the defined scoring rules.

Monitor and Maintain Handscoring Quality

Throughout each of our scoring projects, we place high standards of quality at the center of our delivery goals. To meet those goals, we perform internal audits on our processes to be certain we maintain required quality standards.

The Pearson performance scoring operation is ISO (International Organization for Standardization) certified, and our staff continually monitor for internally determined quality metrics. In conjunction with our ISO 9001:2008 certification, Pearson has developed and documented a standard system for addressing the complexities inherent in monitoring and maintaining quality throughout large-scale handscoring projects.

ISO Certified Quality Management System



Our set of processes and tools provides a replicable quality system that strengthens consistency across projects and locations within Pearson, as well as across multiple years and administrations. We will consult with PARCC upon contract award to customize these processes as necessary to meet the needs of the Operational Assessment program. Additional specifications and procedures will be determined and documented in conjunction with PARCC, and will be approved by the Partnership Manager.

We will use a comprehensive system for continually monitoring and maintaining the accuracy of scoring on both group and individual levels for all Operational Assessment scoring sessions. A description of the major components of our quality assurance system follows.

Daily Systematic Review of Handscoring Reports

Reports on scoring quality and reader performance will be vital to monitoring handscoring quality throughout scoring of the Operational Assessment. These reports are generated and updated automatically and are available to Pearson scoring leadership staff at any time via our digital scoring system.

Our reports give daily and cumulative statistics and provide individual and group average agreement percentages. These reports will be shared with the Partnership Manager daily, and Pearson scoring leads will work with team leaders and the Program Manager to implement any necessary intervention strategies.

See the table in the Handscoring Reports section above for a summary of the reports we will use for monitoring PARCC Operational Assessment scoring. We will also conduct daily scoring conference calls with the Partnership Manager during live operational scoring where we will discuss scoring trends and action plans to address those trends.

Systematic Read Behinds

Read behinds, or backreading, is the immediate method of monitoring a reader's performance and therefore an important tool for our handscoring staff. Backreading is completed in conjunction with the statistics provided in reader performance reports and as indicated by handscoring staff, allowing team leaders to target particular readers and areas of concern.

Our digital scoring system supports systematic, real-time backreading, not only on the first scoring day but throughout the scoring window. Handscoring leads and team leaders can retrieve responses scored by a particular reader, responses receiving a particular score point, or a combination of both.

For responses that have been scored by more than one reader, the system can easily retrieve responses where two readers assigned scores with perfect agreement, adjacent agreement, or non-adjacent agreement. These multiple filters can be used to zero in on any particular reader in any of those situations. This unique capability of our image scoring system allows for timely, targeted monitoring of reader accuracy.

> *Our system allows handscoring leads and team leaders to make immediate searches for responses scored by a particular reader, responses receiving a particular score point, or a combination of both.*

All readers are monitored through backreading as indicated by inter-reader reliability, frequency distribution, and validity reports.

Targeted Read Behinds

Backreading can be targeted towards particular readers and particular score points by carefully reviewing the information provided on real-time statistical reports. Targeted read behinds will be conducted based on reader validity performance.

Handscoring leads review inter-rater reliability, frequency distribution, and validity agreement reports each day, identifying any readers showing particular areas of concern. During daily meetings, they will discuss any reader accuracy issues with team leaders, who will then use the digital scoring system's backreading tool to provide constructive feedback and closely monitor those particular readers until acceptable accuracy is demonstrated.

Of particular relevance to this requirement is the targeted backreading tool in our scoring system. In advance of scoring, performance metrics will be configured in the scoring system. When handscoring leads and team leaders access the targeted backreading interface during scoring, reader statistics will be flagged in red (by reader, by metric) to aid staff in quickly identifying problem areas, for example, in reliability, validity, or frequency distribution (i.e., particular score points).

This process allows us to quickly identify the most problematic issues and resolve them appropriately. For instance, readers showing low inter-reader and validity agreement rates, or those showing anomalous frequency distributions, will be given immediate, constructive feedback and monitored closely until sufficient improvement is demonstrated.

Readers who demonstrate through their agreement rates and frequency distributions that they are scoring accurately will continue to be spot-checked as an added confirmation of their accuracy.

Scoring Validity Responses

Our validity mechanism provides an objective and systematic check of accuracy. Validity papers are actual student responses that are chosen by scoring directors and approved by the Partnership Manager as examples that clearly earn certain scores. These true scores will be assigned to validity responses to compare how often readers match them throughout the scoring sessions. The validity pool will include responses encompassing the entire score range for each item, and readers will read and score them blind (unaware they are scoring validity papers rather than live responses).

Validity statistics match each reader's scores against true scores (much as do qualifying sets). This allows supervisory staff to get a more complete sense of individual and group accuracy and trends.

In contrast to more typical processes that involve readers taking entire validity sets once or twice a day, the digital scoring system provides several advantages:

- The validity responses will be interspersed among live responses to each reader at a regular interval throughout the scoring day, so as to track accuracy and scoring trends more comprehensively—we track reader performance throughout the day.
- Since we route validity responses transparently to each computer, the reader's judgment is completely independent and not affected by the knowledge that his or her score is being compared to a pre-assigned score.
- This method also prevents the test anxiety that many employees feel when they know they are being tested. This blind validity process is therefore acknowledged as a more accurate reflection of readers' true tendencies.
- Because we constantly compile new validity results throughout scoring each day, handscoring leads and team leaders track reader performance more closely and more often.

The validity responses will be interspersed among live responses at the required intervals throughout the scoring day. Our configurable digital scoring system allows us to adjust the validity delivery rates easily, based on direction from the Partnership Manager. The validity insertion rate will be at the rate of one per every 25 actual student responses at the beginning of scoring, and then at the rate of one per every 40 upon direction of the Partnership Manager.

After the first two weeks of scoring, or more regularly as directed by Partnership Manager staff, validity responses may be removed from the set and replaced with new validity responses (to prevent readers from seeing the same responses). Partnership Manager staff may choose to insert new validity responses into the validity set at any time.

The digital scoring system will automatically generate a report that compares the scores given by individual readers with the pre-assigned validity scores. We will use this report to monitor the accuracy of individual readers and the room as a whole.

If a reader drops below an acceptable percentage of accuracy, that reader may require individual feedback and/or retraining before being allowed to score any more responses on the given item.

Automatic Targeting

Given the size and scope of the PARCC program, automatic targeting will be key to managing reader performance across the large scoring pool. Readers performing below 70 percent perfect agreement on validity will receive intervention based on parameters set up in the system and approved by the Partnership Manager in advance of scoring.

Pearson has invested heavily in automation tools built directly in our digital scoring system. We will demonstrate these features for the Partnership Manager and collaborate on how we will apply these automatic tools for PARCC scoring.

The primary tool is called reader exception processing, which allows our project managers to define intervals at which our scoring system checks reader validity or other metrics for exact and adjacent agreement. If readers fall below pre-set standards (after a set interval), messages are automatically sent, interrupting their scoring process. This is an early alert mechanism notifying readers to work with a team leader, review anchor papers, or work though other activities to improve their scoring.

Readers who continue to perform below project standards will receive retraining in the form of targeted calibration. Before taking the targeted calibration set, each reader will be expected to review the training materials, paying special attention to the anchor papers and scoring rubric. The reader will be allowed the opportunity to phone a team leader to ask questions and receive retraining. If the reader does not perform adequately on the targeted calibration set or if the reader continues to perform below the established standards even after completing targeted calibration, that individual will be removed from PARCC Operational

Assessment scoring, and his or her scoring work will be reset and redistributed to other readers.

Targeted Validity (Calibration) Administration

We will create targeted calibration sets (recalibration) to be used as training or remediation for readers whose quality metrics have fallen below the standard. These sets will consist of 10 responses per set. Targeted sets are administered automatically to individual readers whose performance falls below project standards, as described above.

Additionally, our scoring system offers a validity review capability. Within validity review, select validity responses are annotated by the team leaders and flagged (within the system and unknown to the reader). If a reader scores one of these responses incorrectly, the scoring session is interrupted while the response reappears on the reader's screen with the true score, the score he or she assigned, and an annotation. This immediate feedback greatly aids in preventing reader drift. Once a reader has received feedback about a specific validity response, the response is retired for the reader so he or she does not receive it again.

Pseudoscoring

Pearson will not conduct pseudoscoring or require readers to retake all training materials in the event of absences or low quality statistics. Readers who miss work will be required to take all calibration sets they have missed during their absence. Readers who perform below quality standards will receive a warning, and if their quality does not improve after a specified number of validity responses, they will be required to pass a targeted calibration set or be released from the project. Even if they pass the targeted calibration set, they will need to maintain quality standards at every successive validity checkpoint or be released from the project. A reader who is released from scoring an item for quality will have all scoring work on that item reset, and the responses will be redistributed to other readers for new scores.

Group Retraining (General Calibration)

Calibration serves the purpose of retraining of readers.

Individual Conferencing

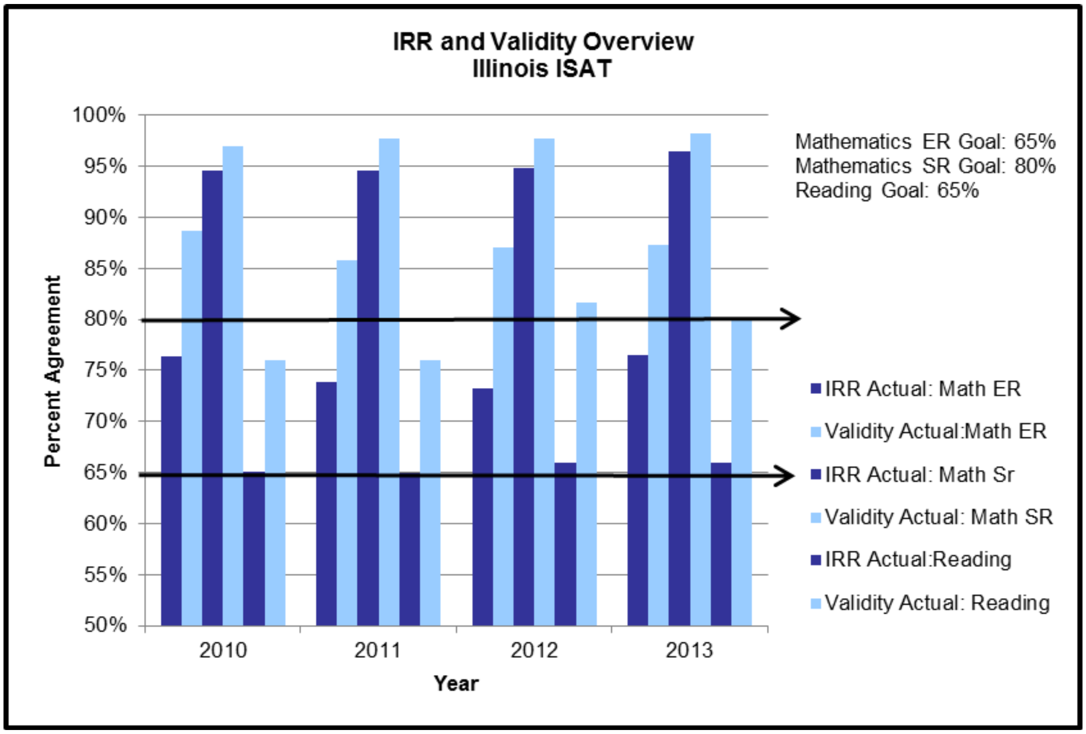
Readers will have access to the dedicated phone, email, and chat tools, which will serve the purpose of scheduled individual conferencing.

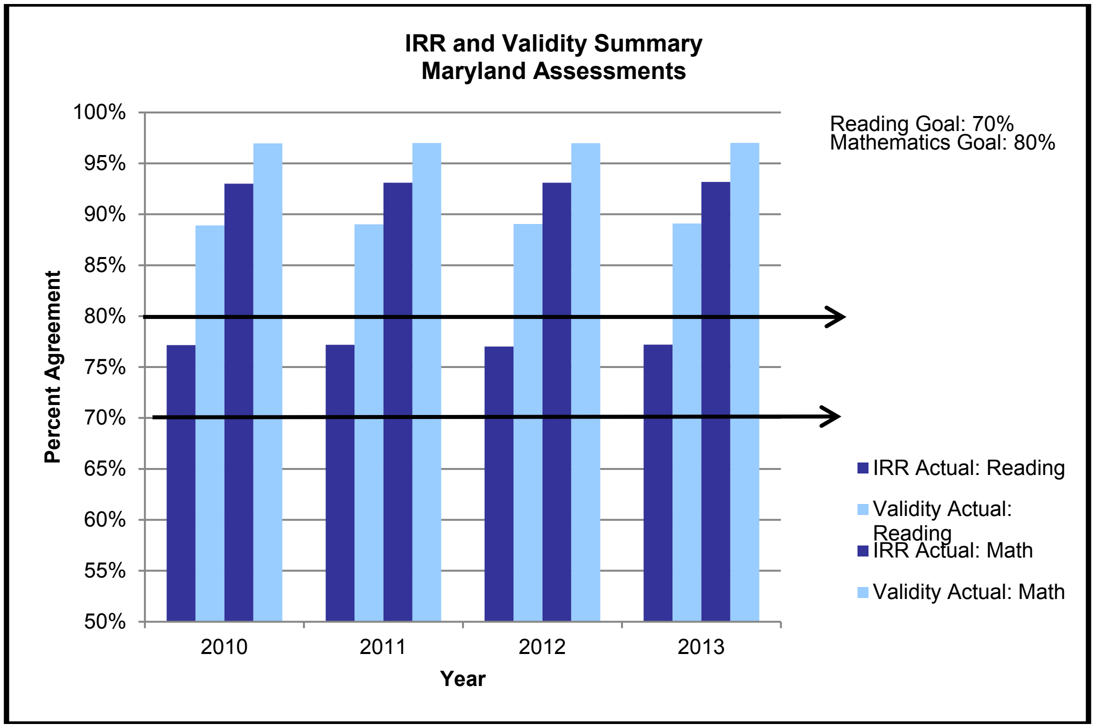
Dismissal

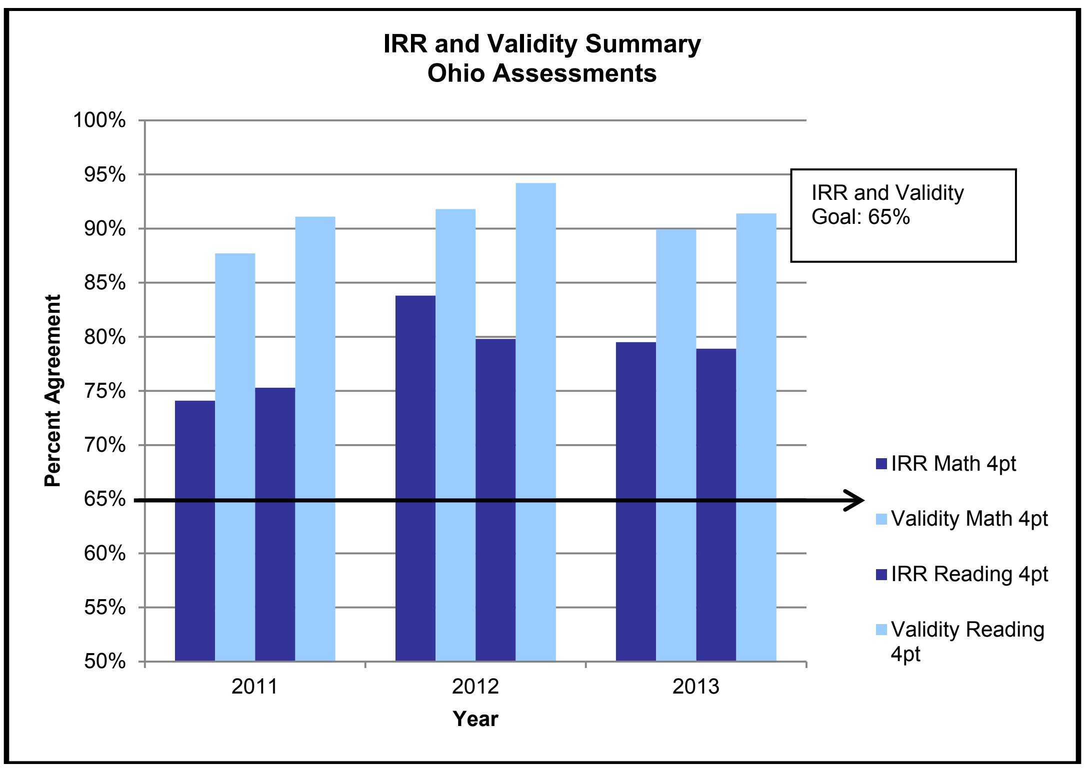
If necessary, our digital scoring system is capable of purging the scores assigned by a reader whose work is deemed substandard. We can reset scores by individual reader, date range, or item. In those cases, the scores assigned by that individual will be cleared from the database and the affected responses will be reset. The responses will then be randomly rerouted to qualified readers and rescored according to the original scoring design.

Evidence of Achieving Quality

Pearson scoring staff members are committed to delivering quality results to PARCC while maintaining efficiency in scoring the Operational Assessment. The figures below illustrate our achievement of quality targets during current contracts with PARCC states. Our prior experience shows that we deliver results and provide accurate, reliable scores.







Consistently High Validity. Pearson has maintained consistently high validity standards on assessments in a number of PARCC states, and will provide high-quality scoring for the Operational Assessment.

Special Scoring Requirements: New York

Special Scoring Requirements for New York are not required under this Agreement, as New York is not a PARCC Participating State for the First Operational Assessment year. In the event New York elects to participate in the PARCC Operational Assessment Administration, Pearson will negotiate directly with New York for any special scoring requirements.

Scoring Specifications Deliverables

V.B.2.K Scoring a. and b.

For items that require hand scoring and do not count toward students' summative scores, it is not necessary to score all student responses. This is the case for embedded field test items, as well as for external anchor items. In both cases, as described in the RFP Section V.B.K.a. and b., we will hand score a representative sample of 1,500 student responses. Test forms will be spiraled at the student level to obtain the most representative sample possible for each item. As such, far more students will be administered each test form than is needed for item calibration and scaling purposes of hand scored items.

Student demographics and the characteristics of the states which comprise the PARCC consortium are relatively diverse. Therefore, to confirm adequate representation, post-stratification will be used to identify a targeted sample of 1,500 valid student responses for hand scoring each embedded field test and external anchor constructed-response item. In some cases it will be necessary to oversample in order to obtain 1,500 valid responses, particularly for items that are very difficult or very easy.

Proper execution of a robust sampling plan forms a solid foundation for analysis. We understand the critical impact representative sampling has on calibration, scaling and equating, and bring that understanding to our work on the PARCC Summative Assessments.

ETS is currently providing sampling plans for PARCC field testing and for the Smarter Balanced consortium, and has done so for a number of K–12 clients (e.g., Virginia, New Jersey, and California). Pearson and ETS will work with PARCC and your TAC to make sure that the sampling plans for hand scored items will yield the most accurate results possible, in the most efficient way possible.

Monitoring and Alerting

Requirement

Response Requirements for Section V.B.2.L.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.L.

Response

Pearson understands that the end user's experience is the most important measure of an assessment system. Delivering at the scale planned by PARCC requires significant risk mitigation, activity monitoring, and customer support.

Mitigating Test Administration Risks

Many test administration issues and testing irregularities can be avoided. Pearson will mitigate foreseeable issues through robust and accessible training. Critical success factors include technical readiness of testing sites, proctor preparedness, and student familiarity with the testing platform via practice opportunities. Pearson will strongly encourage participation in all available training and will provide ongoing support to promote technology readiness.

Monitoring and Support

Pearson has dedicated teams providing all service desk activities, including monitoring, alerting, and escalating. We have made significant investments in trained professionals and tools to monitor and support large-scale assessment activities. Pearson’s Customer Service Center utilizes software to track each incident through to resolution.

While treating each customer contact with the individual urgency required, Pearson is also monitoring incident reports on a larger scale, proactively looking for patterns that may indicate a potential area of concern. To provide a quick response and accurate solution, any alert requiring technical expertise will be escalated to our Level 2 engineers. If further escalation or corrective action is required, the incident will be immediately forwarded to Pearson’s Level 3 engineering team.

In addition to monitoring Customer Service Center contacts, Pearson will also collaborate with the vendor(s) for PARCC’s Technology Bundle and Technology Operations RFPs in order to effectively support the customer-facing activities which utilize the Partnership-owned technology components. Pearson will work with the providers of these components to gain insight into performance monitoring and alerting systems and determine methods for appropriate and timely receipt of performance anomaly alerts.

Resolution and Communication

Responsive customer service is essential in the delivery of high-stakes assessments. Pearson program teams and customer service teams have experience in resolving day-to-day issues that can arise during an assessment project. If an incident is reported by a customer or by a technology component vendor, Pearson will quickly assess the impact and determine an appropriate solution. Our customers can expect a timely response and resolution to isolated issues.

If an uncommon issue arises that could potentially affect PARCC, our First Response Process will guide decision making and provide proper guidance to Pearson personnel regarding escalation to other authorities. The process calls for issue analysis, evaluation of corrective actions by a first response team, and detailed tracking of an issue until its closure. This process will keep both PARCC and Pearson executive management aware of the issue and the progress toward resolution.

In the event of a more widespread issue, or one that is beyond our control, Pearson will implement protocols from an agreed-upon communication plan. Contacts will be made at the appropriate levels, starting with the Partnership Manager and potentially extending to states, districts, and schools. Should the issue require an explanation to the press, Pearson's legal and public relations teams will consult with PARCC's appointed legal representatives to draft an accurate account of the incident, its impact, and the steps to remedy. Pearson will also involve representation from technology vendors as applicable.

Identifying Improvement Opportunities

Effective mitigation, monitoring, and communications are essential, but an assessment of this scale calls for ongoing evaluation and continuous improvement. Through careful analysis of data captured from customer support tickets and technology monitoring systems, Pearson will continue to assess the efficacy of our training content and preparedness strategies. As Pearson invites PARCC to be our partner in this endeavor, we will provide data around incidents and irregularities that is consumable, relevant, and actionable.

Requirement

V.B.2.M. Customer Support and Help Desk Services

1. Customer Support Level 1
2. Customer Support Level 2
3. Customer Support Level 3

Response Requirements for Section V.B.2.M.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.M.

Response

Providing Responsive Customer Service

PARCC member schools and districts will receive customer support and help desk services from the experienced Customer Service Center (CSC) staff at Pearson. The CSC has a strong history of support for diverse state and national assessment testing customers. Formed in 2003, the CSC services approximately 110,000 contacts per year and currently supports 58 individual assessment programs and six unique platforms/products across multiple states.

We will add resources to handle additional call volumes as needed in 2014–2018, although some of the states that we currently support will be transitioning from their current state assessment program to the PARCC Assessment, which will offset some of the additional call volume that will be anticipated with a program of this size.

The CSC has experienced staff members available to respond to the needs of PARCC member schools and districts with questions regarding the PARCC Assessments. Our team members are experienced and competent customer service professionals who can answer questions efficiently and thoroughly. To date in 2013, customers who completed the customer satisfaction survey indicated the CSC met or exceeded their expectations 95 percent of the time.

The Pearson CSC offers a range of services via a toll free number for authorized users, including assisting with implementation, administrative and technical difficulties. Schools and districts can also contact CSS through other methods, as described throughout this section. Below are the levels of help desk support that will be provided to PARCC schools and districts.

- **Level 1 Support**. PARCC member schools and educators will have access to live support via email, chat, and toll-free phone from 7:00 a.m. Eastern time to 6:00 p.m. Pacific time, each weekday, excluding major holidays. This support includes general inquiries, technical and non-technical questions, password recovery, website navigation assistance, procedural and “how-to” questions, etc.
- **Level 2 Support**. PARCC member schools and educators will have access to support for issues not resolved by Level 1 support. These issues may include queries such as how to setup a local system, or other technical related questions.
- **Level 3 Support**. Specialists with detailed program information and focused expertise in the database, network, infrastructure, and software components of web-based services will respond to questions that cannot be resolved at with Level 1 and 2 support. As needed, secure remote desktop support may be used to service issues.

The CSC will also provide technical documentation and appropriate training materials to assist in the resolution of identified issues.

Calls Routed to Appropriate Responder

Using incoming call routing to manage call flow and expedite service, we direct calls to team members best able to answer the question, including specialists to whom calls may be escalated. We will take responsibility for appropriately routing Level 1 and Level 2 calls that require escalation to the next level, without the caller needing to make multiple phone calls. The figure on the following page illustrates the call routing and resolution process.

Robust Knowledge Base Available to All Users

Schools and districts will have **access to self-service using the same knowledge base** used by the CSC staff to obtain information needed to answer questions and provide information to customers. The knowledge base is continuously updated in collaboration with the program team and available for use by all callers.

Trained Customer Service Staff

CSC personnel are trained to quickly determine the scope and impact of reported problems and efficiently route and resolve them. Training for new CSC staff includes the following:

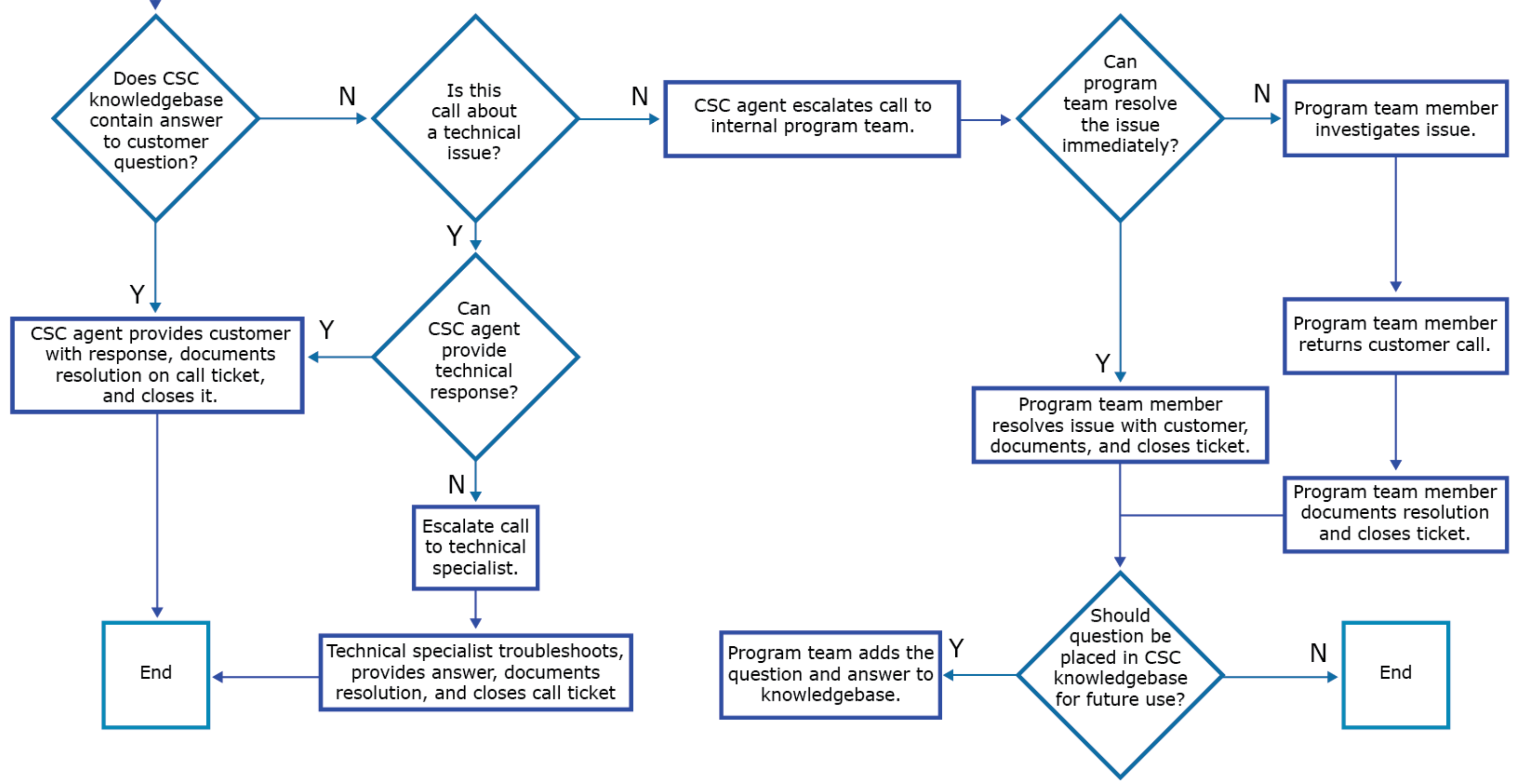
- Overview of knowledge procedures and processes used for supporting our customers
- Pairing up with an experienced staff member to begin listening to customer calls and receiving instruction on phone system operation
- Detailed introduction to and instructions for use of specific websites, products and platforms
- Familiarization with tools used to house knowledge and call information, including Kaidara Advisor software and Peregrine Service Center call tracking system
- Role playing with other staff members and practice with specific call scenarios that may be encountered
- Live customer responses with mentoring by a staff member

Staff members receive regular feedback from the supervisory staff and also receive individual call statistics, so they can see their own progress.

Capacity for Efficient and Dependable Service

Our flexible, scalable operations allow us to provide consistent, quality service during peak periods by routing calls as needed. Operating across a network of centers in Iowa, Minnesota, and Texas, our CSC is more efficient because it is independent of a single location. We create a contingency plan for peak periods using scheduling and forecasting information from the program team and CSC data. In periods of anticipated peak call volumes, we add trained personnel who can log in to the phone system and begin taking calls. This built-in additional capacity will reduce wait time for PARCC callers during peak volumes.

Call Routing and Resolution Process



Call Routing and Resolution. Our escalation protocol will efficiently route ______callers to the appropriate personnel.

Customer Service Systems and Tools

The Pearson CSC offers the following systems and tools to effectively serve PARCC callers:

- **Telephony.** By effectively creating options from which the caller chooses, we use the Avaya telephony system (Avaya G3R Version: V15 w/CM 5.2) to route calls to CSC members trained to handle specific types of issues. This routing allows for a more efficient flow of calls to CSC staff who have the best opportunity to resolve issues during the first call.
- **Chat.** The CSC team uses LivePerson to provide real time customer support. This allows customers the flexibility to interact with agents online.
- **Email.** Emails customers send to the CSC are received by the Service Manager and a ticket number is assigned to the issue. For email received during off hours, we will promptly respond during the next regular business day. Requests for support may also be submitted via a web-based help request form.
- **Knowledge Base.** The CSC team uses the Servigistics knowledge base (Kaidara Advisor 4.4) to access the information needed to answer PARCC caller questions allowing for consistent responses across team members. Because our knowledge base is continuously updated in collaboration with the program team and, as needed, approved by the PARCC, information that may change is promptly available to CSC staff for use in answering questions.
- **Customer Supports.** The PARCC Program Team will maintain materials such as user documentation, training materials and FAQs for the registration and administration processes, and provide to PARCC.
- **Incident Tracking.** All contacts are tracked by the HP Service Manager Client 7.11.259, a secure incident-tracking software tool tailored specifically for the Pearson CSC. With each incoming contact, CSC staff generates a ticket that enables tracking the incident through to resolution, including customer name, date, time and resolution. Tickets are archived and ticket numbers are also provided to the caller. Tickets may be escalated to specialists for their documentation when handling calls to generate a full history of each issue.
- **Quality Monitoring-** Verint Impact 360 Work Force Optimization Suite (v. 11) is used to provide the highest quality service to PARCC. Through recording/monitoring of calls, CSC team members are provided regular feedback from CSC supervisory staff and also receive individual call statistics, so they can see their own progress.
- **Caller Satisfaction Surveys.** To be certain we are meeting your needs, we email surveys to gain valuable information on how PARCC callers view their experience with CSC staff and our response to your concerns.

Log of Received Calls

CSC maintains a record of all calls received including the issue reported, the resolution for the reported issue and the date/time of the call along with the customer name. Calls placed by PARCC member schools and educators can be made available for PARCC representative review.

Summary Reports

Pearson will provide reports to PARCC, at agreed upon intervals, including contact types and resolutions, with customer name, date, and time provided. Also available will be charts for states and districts down to the school level indicating most frequently asked questions. Other reports may be made available upon request.

Requirement

V.B.2.N. Disposition of Materials

Response Requirements for Section V.B.2.N.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.N.

Response

Pearson will continue to protect the security of PARCC data and test content well after tests have been scored and results reported. Paper-based test materials with secure content will be transported, stored, and eventually destroyed according to processes that protect student privacy and valuable test content.

Once data collection and editing processes have been conducted, returned test materials containing student responses will be strapped, labeled, numbered, and palletized for secure storage. Next, they will be transported in locked Pearson trucks to our secure storage facilities that are designed to provide efficient storage of paper-based test materials.

Our inventory and warehouse management system provides end-to-end inventory and tracking of stored materials. Once materials arrive at the warehouse, Pearson staff will use hand-held scanners to catalog materials directly into the system. The location of the test materials will be recorded as they are moved within the facility. This approach will allow the system to track and regulate the movement of stored materials, and it will provide Pearson staff with detailed pallet and inventory reports.

Our approach includes measures for security, materials monitoring, and efficient space allocation. Together, these tools and processes allow us to maintain security, facilitate storage capacity, retrieve stored documents, and stage materials for disposal at the proper time.

The following figure shows our planned timeframes for the storage of test materials.

PARCC Storage Timeframes	
Material Type	**Length of Storage**
Raw materials (unused test material inventory)	6 months following test administration
Secure test books	6 months following security resolution
Documents containing student responses	2 years following scoring and reporting
Electronic Files and print copies	12 months following scoring and reporting
100 copies of each subject/grade test book and answer document	18 months following scoring and reporting
Student response files	2 years following scoring and reporting

Materials will be staged for destruction at the end of the stated time periods. Pearson will seek written approval for destruction in advance. Our costs reflect these timeframes. If PARCC would prefer another storage plan, we can discuss these preferences and the associated costs upon contract award.

Disposal of Secure Test Materials

Pearson moves all test materials through a single, secure disposal path to mitigate risk. Our local recycler will transport the materials to their secure recycling facility in a locked truck.

Our contracts with local recyclers define confidentiality clauses, including their enforcement. Our vendors have current security and alarm systems in place at their facilities to keep unauthorized persons from restricted work areas and their employees sign confidentiality agreements. Pearson also conducts random audits to confirm adherence to our security provisions.

At the recycling facility, materials are securely destroyed, so student data and test items and content are not compromised. As a final precaution, the material is shipped in sealed containers to a paper mill.

Requirement

V.B.2.O. Quality of Work Products

<u>Response Requirements for Section V.B.2.O.</u>

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.O.

Response

Pearson will correct errors in work products at Pearson's expense to the extent such errors are caused by Pearson.

Requirement

V.B.2.P. Quality of Test Administration Study

Response Requirements for Section V.B.2.P.

a) Offeror's proposal shall include a response to the requirements specified in Section V.B.2.P

Response

Pearson will extend the current quality of test administration study beginning with the initial operational administrations in 2014-2015. We will use both the Test Administration Observational Checklist and the Test Administrator Questionnaire for this work.

Based on findings from the initial study carried out in conjunction with the field test, Pearson will recommend any improvements to the two instruments, the supporting protocols for using them, and/or the training processes as appropriate prior to their use in the operational administration.

We will continue to recommend refinements for these processes, as warranted, throughout the contract cycle, although we expect there will be a desire to keep the instruments as stable as possible over time to support longitudinal interpretations of the resulting data.

A critical consideration in using the quality of test administration instruments is the sampling of schools each year and, once a school is selected, the sampling of administrators within a school. The RFP states that observations should be made at 10 schools within each PARCC state. Pearson will work with PARCC and their technical advisors to agree upon a matrix of sampling criteria that will support a representative sampling of the 10 schools in each state.

For example, the matrix might recognize three grade bands (elementary, middle school/junior high, high school), 3 school sizes (small, medium, large), and 3 geographical categories (urban, suburban, rural). Crossing these three criteria would result in 27 possible cells (3 X 3 X 3), and since only 10 schools would be selected, the observed cell proportions of schools in a given state could be used to derive selection probabilities for the cells to guide a sampling effort.

If desired, a particular sampling criterion could be used as a stratification variable. For example, it might be decided that five elementary grade schools, three middle schools, and two high schools would be selected, in which case the conditional cell proportions would determine the selection probabilities for the sampling effort.

Once schools are sampled, another important consideration is what administrations to observe. Pearson proposes to divide the sample such that the performance-based assessments (PBAs) are observed in five of the sampled schools and end of year (EOY) assessments are observed in the other five sampled schools. Within a sampled school, Pearson proposes to utilize the Test Administration Observational Checklist and the Test Administrator Questionnaire and to observe both ELA/literacy and mathematics. For costing purposes, we assume that observers traveling to each selected school will observe the two administrations on consecutive days.

Pearson assumes that for the Test Administrator Observational Checklist, observers will record whether specific administration procedures were followed correctly and successfully, any exceptions to planned administrations, and comments regarding problems or notable best practices.

We acknowledge that the Test Administrator Questionnaire will record basic facts about the administration as well as any problems encountered and how they were addressed. We further assume that both of these instruments will be designed to facilitate aggregated reporting across observations, and we will plan to aggregate the resulting data across schools within a state as well as across states. We will summarize the data collected from both instruments in a report to be submitted to PARCC by August 31 of each operational year.

V.B.3. Automated Scoring

Requirement

V.B.3.A Purpose

Response

Demand for high quality automated artificial intelligence scoring technologies is central to the goals of the Common Core State Standards (CCSS), which mandate more authentic measurement of critical-thinking skills than has been possible with traditional assessments.

For more than a decade, the Knowledge Technologies group of Pearson has been a leader in artificial intelligence scoring by providing visionary technology spanning written and spoken language to assess the traditional four language skills: reading, writing, speaking, and listening. Our scoring technologies now include math, as well.

We have patented applications of cognitive science, computational linguistics, and speech processing technology. Autoscored reading and writing applications include Pearson's unique implementation of Latent Semantic Analysis (LSA). This groundbreaking language modeling technology, along with other automated measures, offers proven validity and reliability scoring student essays—as reliable as professional human readers, and more predictive of the average of two human readers than the inter-rater reliability.

Pearson’s automated scoring engine, the Intelligent Essay Assessor (IEA), has been used to score millions of constructed responses and performance tasks in primary, secondary, and post-secondary education applications, as well as for governments and for publishers and other corporations. One example of its use is on the Council for Aid to Education’s Collegiate

Learning Assessment, for which IEA scores performance task items such as the example below with reliabilities meeting or exceeding human scoring.

> *You advise Pat Williams, the president of DynaTech, a company that makes precision electronic instruments and navigation equipment. Sally Evans, a member of DynaTech's sale force, recommends that DynaTech buy a small private plane (a SwiftAir 235) that she and other members of the sales force could use to visit customers. Pat was about to approve the purchase when there was an accident involving a SwiftAir 235.*
>
> **Document Library**
>
> Newspaper article about the accident
>
> Federal Accident Report on in-flight breakups in single engine planes
>
> Internal Correspondence (Pat's email to you and Sally's e-mail to Pat)
>
> Charts relating to SwiftAir's performance characteristics
>
> Excerpt from magazine article comparing SwiftAir 235 to similar planes
>
> Pictures and descriptions of Swiftair Models 180 and 235
>
> **Using the documents provided, please write a memo to Pat Williams addressing the following questions:**
>
> (1) Do the available data tend to support or refute the claim that the type of wing on the SwiftAir 235 leads to more in-flight breakups?
>
> (2) What is the basis for your conclusion?
>
> (3) What other factors might have contributed to the accident and should be taken into account?
>
> (4) What is your preliminary recommendation about whether or not DynaTech should buy the plane and what is the basis for this recommendation?

Example Collegiate Learning Assessment performance task

Requirement

V.B.3.B. Requirements

Response Requirements for Section V.B.3.

a) Proposed scoring engine
b) Description of proof of concept study
c) A cost proposal that details cost savings the Partnership would realize by using automated scoring, beginning with the first operational assessment. The estimates shall indicate assumptions related the purpose(s) automated scoring will serve (e.g. primary score and/or read behind score) and, for each use, the proportion of responses that will be scored by the automated engine and human scorers.

Deliverables for Section V.B.3.

a) Proof-of concept research design
b) Report of results of proof-of-concept study

Response

Automated, or artificial intelligence, scoring systems promote equity, enable accurate trend analysis, and provide consistent, comparable results for use at the classroom, school, district, or state level. Automated scoring of constructed response (CR) items has grown rapidly in large-scale testing because systems can produce scores more reliably and quickly and at a lower cost than human scoring (see Topol, Olson, & Roeber, 2011; Williamson et al., 2010).

Pearson proposes to engage in the requested proof-of-concept study using both our Intelligent Essay Assessor (IEA) and ETS's e-Rater automated scoring systems. Since 1998, IEA has scored millions of constructed responses administered online to students in grades 4 and above. Some examples of current assessments include the South Dakota writing assessment in grades 5, 7, and 10, the Maryland science assessment in grades 5 and 8, and post-secondary assessments such as the Pearson Test of English and ACCUPLACER®.

Similarly, the e-rater engine has been successfully used for over a decade in a variety of summative high-stakes assessments, most notably TOEFL and GRE, where millions of responses are scored annually from all over the world. Its associated scoring technology has also been successfully deployed in formative low-stakes learning environments through the Criterion companion suite, which provides diagnostic error feedback and subscores that characterize learners in terms of different foundational skills related to writing performance.

Both the Intelligent Essay Assessor and e-Rater use machine-learning approaches in which the scoring engines are trained to score based on the collective wisdom of trained human scorers. Using a representative sample of responses that are double-scored by humans, the computer compares the qualities of each student response to the scores given to the responses by the human scorers. From these comparisons, a prompt-specific algorithm is derived to predict the scores that the same scorers would assign to new responses.

The Intelligent Essay Assessor evaluates the structure, style, and content of writing using a range of artificial intelligence-based technologies. One key technology is Pearson's unique implementation of Latent Semantic Analysis (LSA), an approach that identifies the semantic similarity of words and passages by analyzing large bodies of relevant text. LSA can then be used to represent the meaning of a given piece of text in relation to others, reflecting techniques used by human readers. LSA's representation of meaning allows us to evaluate not only writing itself, but also to measure a writer's demonstration of subject mastery in content areas such as science, history, and social studies.

Proven Validity and Reliability

In tests covering thousands of essays, Pearson's automated scoring technology has proven as reliable as professional human readers and more predictive of the average of two human readers than traditional inter-reader reliability measures. IEA goes beyond simply measuring the grammatical correctness of a response to evaluating its content and completeness. Our automated scoring technology can be used to evaluate written responses in any subject area with reliabilities equivalent to that of human scorers, as shown in the following figure.

Automated Scoring Performance			
Data Sets	**Number of prompts studied**	**Machine-Human Reliability**	**Human-Human Score Reliability**
Prentice Hall LA (6-12)	81	0.89	0.86
MetaMetrics	18	0.91	0.91
Higher Education	8	0.86	0.83

Automated Scoring Performance. Automated scoring can be used in many knowledge-based tasks and yields results that equal or, in some instances, surpass human scoring performance. In addition, automated scoring is unbiased and nearly instantaneous.

A detailed description and explanation of validity testing using Pearson's automated scoring technology is available in a report on the web resource http://images.pearsonassessments.com/images/tmrs/PearsonsAutomatedScoringofWritingSpeakingandMathematics.pdf

Additional information can also be found in Pearson's response to PARCC's RFI on automated scoring submitted in November 2011.

e-Rater has also been shown to be highly reliable, often agreeing with human ratings more strongly than human raters agree among themselves. Additional information about e-Rater can be found at the web resource http://www.ets.org/erater/about.

The combination of IEA and e-Rater affords the Partnership a tremendous depth of knowledge and experience in developing and delivering automated scoring.

Description of Proof-of-Concept Plan

The Partnership has indicated that the proof of concept study shall include Prose Constructed Responses (PCRs) for all three tasks types from the Performance-Based Assessment (research simulation, literary analysis, and narrative writing tasks) and include spring 2014 field test responses from students in grades 3, 5, 7, 9, and 11.

Pearson and ETS propose to study five examples of each of the three PCR types at each of the five grade levels, for a total of 75 prompts. Including five prompts of each type will provide a good sample of data from which to determine the efficacy of scoring PARCC PCRs automatically. The table below provides a summary of the data to be included in the study.

	Task Type		
Grade	**Research**	**Literary**	**Narrative**
3	5	5	5
5	5	5	5
7	5	5	5
9	5	5	5
11	5	5	5
Total per type	25	25	25
Total prompts	75		
Total trait scores	50 x 3 (Literary + Research) + 25 x 2 (Narrative) = 200		

As noted in the table, the research and literary PCRs are evaluated for three different traits (Reading Comprehension, Written Expression, and Language Conventions), while the narrative prompts are evaluated for two (Written Expression and Language Conventions). An automated scoring model will be created for each trait of each prompt, resulting in 200 scoring models overall.

Evaluating automated scoring capabilities involves three phases:

1. Model training/building
2. Model evaluation
3. Model testing

A set of responses will be randomly selected for phases 1 and 2 with the remainder being used for testing in phase 3. Each of these phases is described in detail below. Pearson and ETS will work with Partnership representatives and the PARCC TAC to evaluate and refine the analyses and criteria necessary to fully evaluate the scoring engines.

As part of the spring 2014 field test, we expect to have 1,200–2,400 student responses per regular form and about 250 responses per accommodated form available for model building and evaluation. A randomly selected 20 percent of the responses are planned to be doubly human scored as part of the existing field test contract.

Double-scored responses give the scoring engine a better representation of the full range of scores, including cases in which human scorers might disagree, such as at the line between two score points. Double-scored responses also provide a means of comparison whereby agreement between the automated scoring engine and human scorers can be compared with agreement between two humans.

Because 20 percent double scoring for the field test results in a smaller number of double-scored responses than desired for best practices in training and evaluating automated scoring, as part of the study we will also double human score more of the field test responses to bring the total number of double-scored responses per prompt to 500.

Model Training / Building

Pearson and ETS will train and evaluate their models on the same set of responses. The Partnership has indicated that the study shall address the reliability of the proposed scoring engine when scoring responses written by English learners and students with disabilities, and so the sample used for engine training will be drawn from across the standard and accommodated forms/subpopulations. Not differentiating between standard and accommodated subpopulations for automated scoring model training is also consistent with human scorer training.

Model Evaluation

Model evaluation will include performance on a combined (standard and accommodated) randomly selected sample, as well as on the standard and accommodated subsamples separately. Evaluation criteria for the models are based on criteria advocated by Williamson, Xi, and Breyer (2012), which include measures of inter-rater agreement and external measures.

Measures of inter-rater agreement will include correlations, quadratic-weighted kappas, and standardized mean differences. These measures are computed for pairs of human ratings based on double scoring as well as for pairs of automated and human scores for each engine at the level of analytic trait scores.

We will also compute the measures for each of the two automated scoring engines, as well as for scores representing the combination of both engines' results.

External measures include scores from student performance on state assessments, as well as scores from student performance on non-PCR parts of the PARCC assessment.

Model Testing

Model testing is performed on a set of data randomly selected prior to model building and evaluation phases. This held-out set of data will be scored using the automated scoring models that result from the previous phases. Evaluation criteria for the results will be based on the measures cited above.

The efficacy study will be conducted as soon as possible after the completion of field test scoring in order to have the results available in time to use AI scoring in the first operational year. Below are key milestones:

- Field test scoring complete, including additional double scoring required for study: July 17, 2014
- Scoring data transmitted to ETS and Pearson automated scoring group: July 28-31
- Pearson and ETS independently perform efficacy study: August 1-29
- Pearson and ETS collaborate on draft report and provide to PARCC: September 1-October 15
- Pearson and PARCC agree to continue with AI scoring phase-in plan for year one: October 31
- States that do not intend to use AI scoring in year one opt out: November 14 (add-on pricing applies for ELA scoring for online testing)

Phase-In Plan

Although the efficacy of scoring PARCC ELA/Literacy Prose Constructed Responses has yet to be demonstrated, based on our 15-year experience with automated scoring and our knowledge of the PARCC field test items and rubrics, we believe that automated scoring will be successful on these items and yield great cost savings and schedule benefits to the Partnership.

Assuming a successful proof of concept study, Pearson would propose that automated scoring be phased in as shown in the following figure for all online ELA PCR responses.

Year	% of items	1st Score (100%)	2nd Score (10%)	Resolution
2015	100%	Human	Automated	Human
2016	67% 33%	Automated Human	Human Automated	Human Human
2017	100%	Automated	Human	Human

In the first operational year, we would propose to use automated scoring as the second score on all PCRs. Human scorers would provide the first score for 100 percent of the responses; automated scoring would be used for the 10 percent second read; any resolution scoring would be done by human scorers. All responses would be read by a human, and any scoring discrepancies would be resolved by humans. The first year thus affords us the opportunity to use automated scoring, with its associated cost and time savings, while also providing more data to further validate its use. The Fall/Winter Block 2014 includes paper only assessments. AI scoring will first be used for operational scoring beginning with the Spring Traditional 2015 administration.

Assuming continued success with automated scoring in year one, in year two we will use automated scoring as the first score for approximately two-thirds of the items and as the second score for the remaining third. At this stage, some items may still require confirmation of automated scoring performance. For those items, we would use automated scoring as the 10 percent second read, with 100 percent first read scoring done by human scorers. For the other items, we anticipate being able to use automated scoring as the first read, with human scorers providing the 10 percent second read as a check.

In the third year, automated scoring will provide the first read for all responses, with human scorers providing the 10 percent second score. As in the prior years, human scorers will provide any resolution scoring.

V.B.4. Retest Assessment Administration

Requirement

Response Requirements for Section V.B.4.

Response

General Requirements

Included in the base contract, high school students may retest during regularly scheduled test administration windows (Fall/Winter Block, Spring Traditional/ Spring Block). States needing to offer retesting during the summer can contract with Pearson directly for these services beginning with the summer of 2016. Individual states will have the option of requiring students to take both the PBA and EOY components in each content area or only take one component (PBA or EOY).

If a state requires only one component to be taken as a retest, all students within the state would retake the same component. Retesting pricing for high school students is the same as the per test pricing for students take the high school test for the first time, even if a student only needs to take EOY or PBA, and not both components.

Retest Item/Task Development

As noted in the RFP, additional item/task development for retests is not required. The development of items and tasks for retests is factored into the development quantities in Section V.A.1 of the scope of work. For retests during the regular administration window, the forms will be one of the active forms for that administration. These will be intact forms with an unscored field test section, if part of the form design, that are either active forms or ones from a previous administration. If any states negotiate with Pearson to add a summer retest in 2016 or 2017, forms will be recycled forms from previous test administrations.

Retest Forms Development, Review, Preparation, Printing, Distribution, and Disposition

Retests will be administered in both computer-and paper-based modes. Computer-based tests will be delivered using TestNav.

The forms construction, preparation, review, printing distribution, and disposition of all retests forms will follow procedures identified in Section V.B.2. We will augment training materials developed for the operational assessments to meet the needs of the retest administration. One example of a difference is that individual states will have the option of requiring students to take both the PBA and EOY components in a given content area or only one of those components.

Scoring, Analysis and Reporting

The procedures for key-based scoring, rule-based scoring, and handscoring described in Section V.B.2.K will be followed for retest administrations. In years 2-4, the analysis and reporting of retests will follow the processes described in Sections V.C and V.D of the proposal; however results will only be reported at the student and school level. Because retest students will be using the same form as other students during an administration window, there will need to be a demographic or enrollment category to capture this information and designate a score as a retest.

V.B.5 Practice Tests and Release of Items

Requirement

Response Requirements for Section V.B.5

a) Description of the approach and procedures Offeror will use to identify items and tasks to be released

Deliverables for Section V.B.5.

a) Two blueprint sets with associated scoring materials per assessment for release each year

b) 18 assessment modules per assessment each year
c) Permissions and rights clearance for released items.
d) All QTI and APIP code, associated media files, and metadata for each released item will be quality checked by the Contractor for code confom1ance and validation, accessibility, and releasability (e.g., appropriateness, positive item performance history, permissions), and remediate any necessary corrections, updates, or re-reviews that may be necessary.
e) A manual for accessing and using released items and modules
f) A guidance document for administering the assessment modules
g) A computer-based scorer training module
h) A guidance document for developing and interpreting and using score reports
i) Integrate customer support into the help desk services being provided by the Operational Assessment Administration Contractor.
j) Practice tests for ELA/literacy EOY component and mathematics PBA component to be available in fall 2014, compatible with Assessment Administration Contractor's delivery platform1
k) Full practice tests to be available in fall 2015, compatible with PARCC's assessment delivery platform
l) Scoring materials, including answer keys, rubrics, anchor papers, and scoring rules for use with released items and practice tests
m) Computer-based and paper-based tutorial modules for each grade level/course
n) Four computer-based and four paper-based tutorial modules for each content area

Response

PARCC plans to release PARCC assessment items to promote public understanding of PARCC assessments, demonstrate the range of functionalities of technology-enhanced items, support the interpretation of item-level reporting, and fulfill commitment to the U.S. Department of Education. PARCC plans to release modules to provide teachers and administrators a tool designed to generate information about student progress during the school year.

The Partnership plans to release items each year that would not exceed the total item counts of two operational forms, one online and one paper. The released items will be used for a variety of purposes, such as modules and practice tests. In year 1, the Partnership will select and provide Pearson with enough items and tasks to populate up to two unique blueprint sets of PBA for each content area. In addition to the assessment items and stimuli, PARCC will provide all of the supporting materials such as rubrics, scoring keys and rules, and benchmarked student work. Pearson will select the items for the EOY for all years and the PBA beginning in year 2 according to the parameters mutually agreed upon with PARCC.

Pearson will select items that represent the breadth of the assessment as specified in the test blueprint without jeopardizing the health of the pool for future operational assessments. The items will have a range of content and cognitive complexity. There are several approaches to select items for release. Pearson will work with PARCC to determine the criteria that will be used and understands that it may change as the program matures. Criteria we typically use include the following:

- Items **with high exposure rates** that are not linking items. These items are ones that are used on multiple forms and typically across multiple years.
- Items with **content and skills alignment** where student performance is lagging – this is typical of newer assessments where there has been a shift in the content assessed at particular grade levels and the instructional practices may not be fully in place
- Items whose **data is older** or there is reason to suspect that student performance would be different. These are not good candidates for future operational tests but would represent the knowledge and skills that would be on an assessment.
- Items that represent **newer ways of assessing the content**. This would include some of the technology-enhanced functionality.
- Items associated with **public domain passages**. Many copyright holders do not give permission for non-password-protected online publication or the cost is prohibitive.

Another criterion we may employ, depending on the testing solution for New York, is to pre-identify the PBA forms that would be available for release and use those in New York. Any linking items would be replaced using the criteria above to fill the blueprint.

Pearson will select the items and compose the modules. Careful attention will be needed to avoid compromising the security of any accommodated forms or items appearing on other operational forms. Of the two sets identified each year, one set will be paper and the other will have technology-enhanced items for online deployment. The two blueprints sets will yield 18 modules per content area per year. Modules will not be created for Integrated Mathematics. Permissions and rights will be sought. Payment of those permissions is included in our bid. At least 30% of the passages are assumed to be public domain and the maximum amount for any release items that require permissions is $4000. Scoring information will be available with each module for all non-technology enhanced items. Items and associated materials will be provided to PARCC in PDF and as xml content packages for posting to a PARCC website or the Partnership Resource Center.

The xml files of the items and associated media files will be checked for QTI and APIP code compliance, along with the metadata. The items will have the accessibility features outlined in Table V.A.1.D.2.a. on page 56 of the RFP. It is assumed that the QTI and APIP code and accessibility features for the items provided from the other contract will be compliant.

Manuals

Pearson will develop and compose up to three manuals to support the released items. We will work with PARCC to determine the purpose, audience, and intended use of each manual. The content for the manuals will be a joint effect between PARCC and Pearson. The manuals will be available in PDF for posting to PARCC's website or the PRC. Possible manuals include

- A manual for accessing and using released items and modules

- A guidance document for administering the assessment modules
- A guidance document for developing and interpreting and using score reports

Once the manuals have been created, the manuals will be reviewed and revised annually.

Scorer training module

Pearson will develop a scorer training module that will instruct teachers on how to print out constructed responses for students that require human scoring and how to score the responses to Type II and Type III mathematics items and PCR ELA/literacy items. The scorer training module will be available to educators as a PDF accessible on PearsonAccess and through the Partnership Resource Center.
An important part of the scorer training modules will be effective instructions for teachers to reliably score student constructed responses that cannot be automatically scored by TestNav. To accomplish this, Pearson will include the rubrics, instructions for using rubrics and areas of focus for each item type, training materials used by human scorers to score the items, and benchmarked student work (anchor papers) for each score point. As part of this module, teachers will also be provided several papers to score and feedback on the score they gave the responses. This will be similar to a qualifying set used with scorers.

Practice tests for ELA/Literacy EOY component and mathematics PBA component to be available in fall 2014

PARCC's vision for the assessment program includes practice tests, enabling both students and teacher to gain valuable experience with the assessment items and tasks. Practice tests represent a critical step in supporting students' and schools' transitions to new assessment content, as well as providing teachers and administrators a tool to generate student performance data to inform instructional improvements and professional development. This experience will also provide students with a hands-on experience with the items and tasks well before the first operational administrations are delivered.

PARCC will produce the practice tests for the PBA component of ELA/literacy and the EOY component for mathematics. These will be available in spring 2014. Practice tests for ELA/literacy EOY component and mathematics PBA component will be available in the 2014-2015 school year, at least six weeks prior to beginning of the spring test administration window on TestNav. High school practice tests will be available in November 2014, the month before the Fall/Winter 2014 Block administration. The items for the practice test will be some of the released items. No new item development is planned.

Pearson has extensive experience in developing practice tests and public release materials for the purpose of familiarizing teachers and students with the variety of item types and tasks that comprise the operational assessments. For each practice test, Pearson will evaluate the test blueprint and select a representative subset from the released items that includes the broad range of task types and functionalities that the student will experience in the operational assessments. Pearson will work with the Partnership to determine and finalize the specific distribution of practice test items and tasks by domains and strands.

The Partnership will provide Pearson with items, tasks, passages, metadata, scoring materials, in the first year (2014-15) after the items and tasks have been field tested. After Field Test analyses are complete and items that are candidates for release have been identified, Pearson will build PARCC practice tests comprised of EOY tasks for each grade level, 3–11, in ELA/literacy for a total of 18 forms assessing ELA/literacy (nine computer-based forms and nine paper-based forms) and 18 forms comprised of PBA tasks assessing Mathematics (9 computer based forms and 9 paper-based forms); one form for each grade. No practice tests will be developed specifically for Integrated Mathematics.

. After the first year, practice tests will be updated as needed, incorporating items with new functionalities as they become available. Practice tests will be available no later than six weeks prior to the first administration of each summative component, as shown below.

Practice Tests						
			2014-15*	**2015-16**	**2016-17**	**2017-18**
PBA	ELA/L	Grade 3		Dec. 2015	Dec. 2016	Dec. 2017
		Grade 4		Dec. 2015	Dec. 2016	Dec. 2017
		Grade 5		Dec. 2015	Dec. 2016	Dec. 2017
		Grade 6		Dec. 2015	Dec. 2016	Dec. 2017
		Grade 7		Dec. 2015	Dec. 2016	Dec. 2017
		Grade 8		Dec. 2015	Dec. 2016	Dec. 2017
		Grade 9		Sept. 2015	Sept. 2016	Sept. 2017
		Grade 10		Sept. 2015	Sept. 2016	Sept. 2017
		Grade 11		Sept. 2015	Sept. 2016	Sept. 2017
	Math	Grade 3	Dec. 2014	Dec. 2015	Dec. 2016	Dec. 2017
		Grade 4	Dec. 2014	Dec. 2015	Dec. 2016	Dec. 2017
		Grade 5	Dec. 2014	Dec. 2015	Dec. 2016	Dec. 2017
		Grade 6	Dec. 2014	Dec. 2015	Dec. 2016	Dec. 2017
		Grade 7	Dec. 2014	Dec. 2015	Dec. 2016	Dec. 2017
		Grade 8	Dec. 2014	Dec. 2015	Dec. 2016	Dec. 2017
		Algebra I	Nov 2014	Sept. 2015	Sept. 2016	Sept. 2017
		Geometry	Nov 2014	Sept. 2015	Sept. 2016	Sept. 2017
		Algebra II	Nov 2014	Sept. 2015	Sept. 2016	Sept. 2017
EOY	ELA/L	Grade 3	Jan. 2015	Jan. 2016	Jan. 2017	Jan. 2018
		Grade 4	Jan. 2015	Jan. 2016	Jan. 2017	Jan. 2018
		Grade 5	Jan. 2015	Jan. 2016	Jan. 2017	Jan. 2018
		Grade 6	Jan. 2015	Jan. 2016	Jan. 2017	Jan. 2018

Practice Tests						
			2014-15*	**2015-16**	**2016-17**	**2017-18**
		Grade 7	Jan. 2015	Jan. 2016	Jan. 2017	Jan. 2018
		Grade 8	Jan. 2015	Jan. 2016	Jan. 2017	Jan. 2018
		Grade 9	Nov 2014	Oct. 2015	Oct. 2016	Oct. 2017
		Grade 10	Nov 2014	Oct. 2015	Oct. 2016	Oct. 2017
		Grade 11	Nov 2014	Oct. 2015	Oct. 2016	Oct. 2017
	Math	Grade 3	[redacted]	Jan. 2016	Jan. 2017	Jan. 2018
		Grade 4		Jan. 2016	Jan. 2017	Jan. 2018
		Grade 5		Jan. 2016	Jan. 2017	Jan. 2018
		Grade 6		Jan. 2016	Jan. 2017	Jan. 2018
		Grade 7		Jan. 2016	Jan. 2017	Jan. 2018
		Grade 8		Jan. 2016	Jan. 2017	Jan. 2018
		Algebra I		Oct. 2015	Oct. 2016	Oct. 2017
		Geometry		Oct. 2015	Oct. 2016	Oct. 2017
		Algebra II		Oct. 2015	Oct. 2016	Oct. 2017

In building the practice tests, Pearson will meet PARCC's goal of constructing forms that provide students in PARCC states with the opportunity to complete a single form of grade-level items in approximately 60 minutes. Pearson will work with PARCC to identify and select items and tasks that give students the best opportunity to experience the types of items on the PARCC assessment.

Since innovation and innovative item types are important elements of the PARCC assessment experience, we will prioritize the selection of practice test items that sample innovations in both ELA/Literacy and Mathematics, provided the item pool will support the release.

Using the approved form specifications, Pearson will build a practice test for each grade/course level and content area so students can complete the each practice test in approximately 60 minutes (as stated in section V.B.5.B) and also representative of the test blueprints. Although close adherence to the blueprints for both content areas is not a specified requirement in the RFP, Pearson sees several advantages in building test forms for public release that resemble the blueprints to the greatest extent possible. These advantages include the following:

- The release to the field of an accurate representation of the test content
- The opportunity to familiarize teachers and students with as many different items types as possible in a small sample of the larger test
- The opportunity for teachers to align instruction and assessment
- A balanced distribution of score points that approximates those generated by the operational assessment

Each practice test will be made available to all students in all PARCC states in both computer-administered mode and paper mode (as a printable PDF from PARCC website). Because the tests will be readily available to students, we assume that the practice tests will not be secure.

Our additional assumptions include the following:

- **Consistent form construction process**. For the practice-test form composition and publication, Pearson will follow the same detailed procedures for computer-based and paper-based test forms described in our response under sections V.B.2.C. We will also follow the same process for key verification described in our response to section V.B.2.K, with the one exception that the practice test will not require a post-administration statistical key check.
- **Consistent Online Delivery Platform.** Practice tests will be delivered using TestNav, the same online assessment platform used for operational administration. . This will help to familiarize teachers and students with the interface and assessment tools. As PARCC stated in its original grant application, “Maintaining the same user interface and design for various assessment components means students and teachers will need little to no instruction to quickly begin the assessments after an initial training period.”

For the Practice Tests delivered on TestNav, machine-scored items will be scored and scored responses will be provided to the student after he/she submits responses to all items on the practice test. As specified in the RFP, those items that require hand scoring will not be scored by the contractor but locally by teachers. To facilitate local scoring, Pearson will provide scoring materials, including rubrics and benchmarked papers.

Full practice tests to be available in fall 2015

Beginning fall 2015 and through the remaining duration of this contract (Spring 2018), Pearson will review and modify practice tests annually to incorporate items with new functionality. These tests will be available not later than six weeks prior to the opening of each administration outlined in Table V.B.2.A (Annual PARCC summative administration windows) of the RFP on TestNav.

Pearson will build PARCC practice tests comprised of EOY tasks for each grade level in ELA/literacy and each grade level/course in mathematics for a total of 36 forms (nine computer-based forms and nine paper-based forms in ELA/literacy and 9 computer-based forms and 9 paper-based forms in mathematics). Pearson will also build the same number of practice test (36) comprised of PBA tasks for both ELA/literacy and mathematics—one computer based test for each subject and grade/course, except Integrated Math I, II, and III and one paper based practice test for each subject and grade/course, except Integrated Math I, II, and III.

Scoring materials, including answer keys, rubrics, anchor papers, and scoring rules for use with released items and practice tests

Pearson will assemble all scoring materials for the released items and practice tests in manner that is clear for teachers and administrators to understand. The answer keys and scoring rules will be the same ones used for scoring the items from when they were last tested and will go through the same key verification as described in section V.B.2.K, with the one exception that the practice test will not require a post-administration statistical key check. Pearson will also include both the anchor papers and rubrics for all human scored item that were used when these items were last tested.

Computer-based and paper-based tutorial modules for each grade level/course

To better prepare students for the summative tests, Pearson will create both computer-based and paper-based tutorial modules to be available for each grade level and course. Pearson will design four tutorials per mode to allow students to become familiar with the testing mode they will be using for their operational tests. The tutorials will be developed to be available by grade bands that have similar item functionality types and tools. One possible configuration is grades 3-5, grades 6-7, grade 8, and HS, based on calculator type.

Each computer-based tutorial will include released items especially selected to allow the student to interface with the embedded tools and supports in PARCC's test delivery system for the corresponding subject, grade, or course level. Pearson will also include as wide of variety of items types possible (corresponding to each subject and grade level or course) while limiting each module to approximately 30 minutes to complete. Computer based tutorials will be compatible with PARCC's test delivery system and also accessible on PARCC's PRC.

The four paper-based modules will be designed for students to use a separate answer document to record their responses just as the students would do for the operational or standalone field tests. Pearson will also select items that will allow students to become familiar with the paper based tools (e.g. rulers in math), item types (e.g. multiple choice, multiple select, gridded response, and constructed response) while limiting the time for students to complete the module to approximately 30 minutes. Paper based modules will be available as printed electronic documents hosted on PARCC's PRC.

Each year, Pearson will update each module based upon any new tools, embedded supports, item type, or navigation as well as feedback from the Partnership. Pearson will work with the Partnership to create a detailed schedule that allows Partnership representatives, including accessibility experts, to review, revise and approve each module to support them being available on PARCC's PRC as least four weeks prior to the first operational assessment each school year—October for high school and January for the other grades.

Alert Systems

Hand-in-hand with delivering accurate scores to PARCC, Pearson is also concerned with the safety and well-being of all students in PARCC member states. We train our scorers to be aware of student responses indicating potential need for intervention.

Alerts include student responses that might indicate that a child is a danger to him or herself or others, is experiencing depression, is involved in abuse, or is contemplating suicide. Other alerts include responses that might indicate a testing irregularity may have occurred. We will follow our standard process for alerting responses that may require intervention by testing or school officials.

Process for Handling Alerts

Scorers receive initial instruction regarding alerts during training. Scorers are instructed to contact their supervisor, even if they are unsure if intervention is required. Our staff does not make that determination, but rather will forward any response in question to the Partnership Management staff for appropriate handling.

When a response is alerted, the scorer includes a brief comment explaining the issue and notifies supervisors. Pearson lead scoring staff reviews the alert and completes a standard alert form, which includes a brief description of the issue and the unique identifying number associated with the response. These documents are forwarded to the Pearson program team, who use the tracking information to link the response to the student record. The complete documentation will then be forwarded to the Partnership Manager for proper handling. Technology allows us to immediately notify PARCC; we do not wait until the end of the project to process alerts.

Pearson takes steps to safeguard that the correct student response is alerted when the student's writing indicates the child is a danger to himself or herself, or to others. If the alert involves response from a pen and paper test that has been scanned, Pearson program team members access the original hardcopy document using the unique response identification number to compare the electronic image to the hardcopy and validate the correct response has been accessed.

Because of the nature of the material and lack of appropriate context, we are unable to determine whether threats or other statements contained in test responses were made seriously or humorously or whether they were intended to be interpreted as real or fiction. The alerted response and documentation is referred to the Partnership Manager so that the local authorities can take the action that they deem appropriate under the circumstances.

Responses will be alerted if they are thought to indicate one or more of the following:

- The student is a danger to himself or herself or to others.
- The response suggests a situation which warrants investigation such as the possibility of abuse, depression, or contemplation of suicide.

If Pearson refers a student's test to PARCC, we will do so without making any evaluation or recommendation, other than to make note of the questionable content.

Other Alerts

In addition to Child in Danger alerts, our electronic scoring system allows other separate review queues to be set up. Scorers can send papers to these queues that indicate the need for further review. Scorers are trained to contact their supervisor concerning responses that indicate that there may have been cheating, or that a teacher answered for the student, translated for the student, or interpreted the student's response. In each case, our scoring directors review the response to document the issue and contact the Partnership Manager as appropriate.

For machine-scored responses, our scoring engine is trained to detect child in danger papers. Responses that do not meet the scoring engine's specifications for scorability are scored by humans.

Disclaimer and Limitation of Liability

Pearson makes no representation that student responses are screened for indications of violence, abuse, neglect, or any other conduct. To the contrary, Pearson specifically states that its educational measurement products are not intended to serve as indicators of violence, abuse, neglect, or other questionable conduct, and that Pearson itself does not have the ability to make any such assessment. Therefore, Pearson policy and procedures should not be relied on by any school district, student, family, or anyone else as a means of detecting or assessing violence, abuse, neglect, or other questionable conduct.

Also, while our scorers all have four-year college degrees, they do not necessarily have training or expertise in psychology or counseling. For these reasons, it is impossible for our scorers to know whether statements made or situations described in student essays are real, exaggerated, or imagined. Nevertheless, because certain statements raise potentially serious concerns, especially where indicating the possibility of violence to self or others, we have instituted policies that call for the reporting of such statements to school authorities.

Pearson will make good faith efforts to follow our alerts policy and these procedures. However, any failure to follow the policy and procedures shall not be deemed a breach of contract or held against Pearson. No claims of any nature may be made by customers against Pearson arising out of Pearson's referral of questionable content or its failure to refer questionable content.

Information to Students and Parents

Because Pearson's policy may in some instances lead to the identification of student responses as sources of questionable content, it may be appropriate for students and parents/guardians to be informed of the policy. The state may want to consider sending a statement to all students and their parents/guardians regarding this policy before any open-ended educational assessments are administered to the student.

Component 4: Reporting

V.D.1. Finalize Design of Reports

Requirement

Response Requirements for Section V.D.1.

a) Detailed plan for finalizing PARCC report designs for all reporting levels, including specifications on how the Offeror would work with the Technology Bundle Contractor to finalize the designs.

b) Detailed plans on how the Offeror would create released and unreleased item descriptions based on the knowledge and skills needed to correctly answer the problem.

Deliverables for Section V.D.1.

a) Final PARCC report designs for year 1 for the following reporting levels:
   i. PARCC Level Report
   ii. State Level Reports (with minimal customizable options)
   iii. District Level Reports (with minimal customizable options)
   iv. School Level Reports (with minimal customizable options)
   v. School Roster Reports (with minimal customizable options)
   vi. Student Level Reports, including specifications for incorporating released items into student level reports (with minimal customizable options)

b) Updated PARCC report designs for years 2 and beyond for all of the above reporting levels (as needed), including translations.

c) Student level data files

d) Summary data files

Response

Pearson will work with the contracted vendor and will use the results and final mockup reports from the Score Report Design Study under the Data Warehouse and reporting development vendor and make any final revisions needed, as determined by the Partnership, to produce the required report designs. We anticipate that the data management and reporting development contractor will have turned over all the necessary training and documentation needed for Pearson to learn and configure the system. After receiving and reading all training on the reporting system, Pearson will work closely with the data warehouse and reporting development vendor to understand the final state in which the reports will be turned over and understand the results of the reporting design studies.

In order to begin the work to finalize the PARCC report designs, we anticipate that the PARCC Technology Data Warehouse and Reporting Development vendor will provide the documentation, mockups, and the results of the study. When the software is made available to us for configuration and management, we expect that the final study templates will have been created in the system and we can make the necessary modifications to match the final report mockups provided by Pearson through this contract. Pearson will be responsible for the quality of the summative assessment data uploaded into the system, including data import, data quality checks and report quality checks for summative assessment data, not including dynamic reporting available through the Data Warehouse.

To understand PARCC's needs for reports and data files, we will use the following processes, to document PARCC requirements and verify that our deliverables meet the Partnership's goals.

Step One: Customer and Internal Requirements

Shortly after award our staff will discuss PARCC's specific reporting requirements. Our requirements analyst will re-examine the mockup reports from the data warehouse and reporting development vendor along with PARCC's requirements to finalize the design to be used. PARCC requirements will direct how data of various types will be handled, such as the following:

- Will incomplete responses to items be given partial credit, and if so, how will this "attemptedness" be represented in test item data and on student reports?
- Will results from test takers who have recently relocated be included in the summary for the new or previous district? How will this data be represented?
- To protect student privacy, what is the minimum number of students in any demographic reporting group? How should data be presented for groups falling beneath this threshold?

PARCC report designs for Year 1 will include the following reporting levels:

- PARCC Level Report
- State Level Reports
- District Level Reports
- School Level Reports
- School Roster Reports
- Student Level Reports, including specifications for incorporating released items into student level reports
- Student Level Data Files
- Summary Data Files

Pearson will translate the Student Level Report into the 10 languages designated by PARCC in the summer of 2014, as described in Section V.A.2 of the RFP. We will prepare PDF versions of each translated report and will provide the versions for posting on PARCC's website and in the PARCC Partnership Resource Center.

Step Two: Score and Report Planning Meetings

Early in the contract cycle, we will schedule scoring and reporting meetings (included in the Project Specific Meetings in the Cost Proposal). During these face-to-face meetings between PARCC, Pearson, and the data warehouse and reporting development vendor, we will verify and update the requirements documentation and report mockups. An agenda item for the meeting will include the plan for having the item descriptions for released and unreleased items on the Student Level Reports. Therefore, it will be important to include not only technologists and report design staff, but also PARCC and Pearson Team content experts in the meeting. For non-released items it will be especially important to understand the level of detail that will be included, in order to be helpful in conveying the knowledge and skills assessed, without the need for the student to cross-reference a number of other materials. Once we understand the level of detail that will be provided, our content team, which includes staff from ETS, WestEd, and Pearson, will draft item descriptions as part of the test construction process. The descriptions will not be created until after items have been selected for an operational test. The descriptions will be at the standard level whenever possible. The descriptions will be reviewed and approved by PARCC.

Step Three: PARCC Confirms Final Requirements

Our program team for PARCC, report designers, and technology developers will continue to work with PARCC to baseline this documentation in anticipation of future scoring, data analysis, and reporting activities.

Prior to each school-year, we will gather the following types of information and verify for PARCC acceptance:

- Project milestones related to reporting
- Confirmation of reporting deliverables
- Final report types and descriptions
- Business rules for each field displayed
- Suppression and exclusion rules
- Identification of student strengths and weaknesses
- Item-level reporting requirements
- Expected collation and pagination
- Media and destination requirements
- PARCC standards and grade-level equivalencies
- Effective real-world graphic displays
- Intended report recipients: PARCC, state administrators, district coordinators, schools, and students

Our customer-requirements analyst and program team leaders will have captured the most significant details about PARCC's intended output. Final sign-off on reporting requirements from PARCC is requested by December 2014. We will also engage in an iterative review process with PARCC to obtain sign-off on the final report mock-ups.

Step Four: Completing the Reporting Requirements

In some cases, this final step requires several weeks to finish, which is why we begin the process so early. Our customer requirements documentation and final mock-ups will be used later during data testing, quality assurance evaluations, and production to validate that our deliverables meet PARCC's reporting needs.

In Year 1, Pearson will collaborate with the Partnership and the data warehouse and reporting development contractor to produce final report designs for PARCC, state, district, school level reports, school rosters, and student level reports. Then, in year two, any necessary updates to the reports will be applied by Pearson and used by Pearson for the remaining years.

V.D.2 Data Upload and Generating Reports

Requirement

Response Requirements for Section V.D.2.

a) Detailed quality analysis plan for initial data upload.
b) Detailed plans for a discrepancy resolution window during which states, districts or schools address data quality issues before final reports are posted.
c) Detailed quality plan for receiving corrections from states (and as needed from districts/schools) prior to reporting.
d) Detailed quality plan to assure that information available in the Partnership's dynamic reports is correct.
e) Plans for determining whether scoring/reporting errors have been discovered and protocols for correcting the error and plans for providing regular reports.

Deliverables for Section V.D.2.

a) Within 30 days of contract execution, the Contractor shall work with PARCC to provide a detailed of all electronic reporting files needed. The Contractor shall be responsible for generating each of these reports and files.
b) Uploaded data files for each summative and retest administration.
c) Detailed reports on when scoring/reporting errors have been discovered and protocols for correcting the error.

Response

For each summative and retest administration, Pearson will upload the scored data file into the PARCC data management and reporting system. Pearson will perform quality checks on the data after upload into the PARCC data warehouse and reporting system to confirm the data warehouse has been populated accurately with uploaded data.

.

Detailed Quality Plans

Prior to uploading the scored data file into the data management and reporting system Pearson will perform quality checks to verify the integrity of the data upload. We will comply with the PARCC's Data Privacy and Security Policy, as recently approved by the Governing Board. The following process describes the key review quality steps that will be followed to produce scored data.

As part of our ISO 9000 standardized best practice test development process, we routinely perform a key review prior to test administration to verify that the scoring (answer) keys are correct for each form. Once the forms have been constructed and are approved by PARCC for publication, an independent, internal key review will be performed by experienced and trained content staff. The content staff will review each item and confirm that the key is correct. If discrepancies are identified, a second senior content specialist or content manager will review the flagged item(s) and work with the item developers to resolve the issue. Our internal key review process will minimize the probability and risk that PARCC will identify errors in the scoring keys or need to make additions to the scoring keys.

Another key review check runs standard code to compare the multiple-choice item keys in the customer test map against what is in the item XML code to check for discrepancies. All discrepancies will be resolved internally and will then be presented to PARCC for approval.

When all key reviews have been completed prior to scoring, Pearson will provide PARCC the final keys that will be used for scoring for PARCC to approve.

We will work with PARCC on any outstanding discrepancies and/or errors and make the necessary updates or revisions in the item bank and resubmit the test key information for PARCC's final approval.

Quality assurance in the form of data validation includes both verification that our systems or processes operate according to functional specifications and validation that the output from a system or process meets PARCC requirements.

The following figure shows the standard sequence of verification steps for scoring and reporting, ultimately yielding the complete data set of reports.

Capability	Standard Validation Activity
Verify items are machine scored correctly	Content experts confirm that items map accurately to the correct response. Scoring keys are verified.

Capability	Standard Validation Activity
Verify handscoring scores are merged with student records	Pre-defined and approved scoring rules are incorporated into the scoring system. Test plans are written to confirm that the correct number of readers have scored each response. Quality engineer confirms the output against expected results.
Verify aggregated scores are correctly reported	The reporting system is configured based on pre-defined and approved scoring specifications. Test plans reflect how the rules for aggregating and rounding are applied to produce summary performance data. Quality engineer confirms the output against expected results.
Verify final data files and reports	Presentation of data on reports of results is validated against pre-defined specifications and approved report mockups. Required data elements and file formats for final data files are verified as defined in the test plan.

Report Data Verification. Established and repeatable Pearson processes will provide data validation at key stages on the way to delivery of final results data for PARCC.

Once the scored data has been uploaded, data warehouse has been populated accurately with uploaded data, states, districts, and schools will have a predetermined time agreed upon between PARCC and Pearson, to review the data for issues before the reports are considered final. Any problems identified during quality review performed by Pearson, states, districts, and/or schools will corrected by Pearson before release of reports. We will provide users with a detailed plan for how to review results for possible quality issues (e.g., reviewing total number of students that tested, number of sites, etc.) and for instructions on how to notify Pearson if there is a potential error once Pearson has been trained on the reporting system. Once the state, district, and school reports are final and have been re-uploaded, as needed, Pearson will follow a similar process with PARCC for validating consortium-level reports.

Reporting Schedule

The following are major milestones for the Year 1 reporting schedule. Without understanding the schedule for the data warehouse and reporting development vendor's report design study, it is difficult to provide a detailed schedule at this time. We will work with PARCC and the Technology Bundle vendor to develop a more detailed schedule, which will include specific and interim dates and handoffs leading up to the required approvals.

Reporting Schedule	
Contract Finalized	April 2014
PARCC provides list of electronic reporting files needed	May 2014
Pearson provides a detailed plan and timeline for delivering student reports (if required) and critical milestones	July 2014
Data management and reporting vendor provides results of report design study and draft report	To be provided by vendor

designs to Pearson	
Reporting requirements approved by PARCC	December 2014
Standard Setting for high school assessments	July 2015
Governing Board approves cut scores for high school assessments	August 2015
Standard Setting for grades 3-8	August 2015
Governing Board approves cut scores for grades 3-8	September 2015
Scored data for 2014-2015 operational assessments uploaded to the data management and reporting system	November 2015

Year 1 Reporting Milestones. We will work with PARCC and the Technology Bundle vendor(s) to develop a more detailed schedule.

We will provide two copies of the individual student report for all students. For students in grades 3-8, the report will be a four page folder and include results for both mathematics and ELA (PBA and EOY), based on mutually agreeable matching rules defined by PARCC and Pearson. For any student demographics that appear not to have a match, mathematics and ELA will be reported separately. Students will test in a single mode (paper-based or computer-based test delivery, unless the state has otherwise made arrangements with Pearson for a specific phase-in plan for computer-based testing. Students in high school will have results for each test (e.g. Algebra I) reported on single sheet, duplex report that includes both PBA and EOY components. For a high school student taking multiple PARCC high school assessments within the same administration window (e.g., ELA I and Algebra I), they will receive separate reports for each test. We will work with PARCC to develop a specific timeline for providing paper reports in Year 1.

The reporting schedule for Year 2 will depend upon a number of variables such as the following:

- PARCC's approval for the use of pre-equating in Years 2–4
- State and district testing windows within the overall administration window
- Actual number of students that participate in each administration, the percentage of students that test online, and the phase-in plan for automated scoring if approved as an option by PARCC
- On-time return of materials from schools and districts—the RFP requirement to allow sites 7 days after testing before materials need to be collected by the vendor will delay getting timely results back to schools and districts. PARCC could require smaller districts and schools to return materials even earlier for scoring.

<u>We can provide electronic copies of student and school level reports without the entire district, state, or consortium having completed and returning testing materials so that schools may receive results prior to the end of the school year.</u>

Prior to year 2, when high school students may retake an assessment as part of a regular testing window, PARCC and Pearson will work together to determine how retest scores will be reported and what data needs captured during registration/pre-ID to allow for this.

V.D.3. Score Interpretation Guides

Requirement

Response Requirements for Section V.D.3.

a) Detailed plan for developing and implementing print-ready and web-based, interactive score interpretation guides for each of the summative and retest assessments that are inclusive of the final-report designs.

b) Detailed plan for user-feedback cycles to assess end-user understanding of the print-ready and web-based interactive scoring guides prior to system-wide release.

Deliverables for Section V.D.3.

a) Separate print-ready and web-based, interactive score interpretation guides for each of the summative and retest assessments that are inclusive of the final-report designs.

b) Materials for a one-hour webinar for parents and educators each for report interpretation.

c) Recording of one-hour webinars for parents and educators each for report interpretation.

d) Annual updates to all print and web-based materials reflecting any changes to the assessment reports.

e) Translated materials to interpret parent reports of up to 10 languages.

Response

Overview

Pearson will develop score interpretation guides for use by local and state education agency staff. We will make these guides available in a PDF print-ready format, as well as an Internet-based version that can be accessed interactively at one or more pre-determined web sites. These guides will include graphics portraying all types of reports, and will include narrative text describing the features and information presented in each report, and showing users how to access the reports. We will include information specific to teachers to help them explain the score reports to parents in ways that can help them better understand their children's levels of knowledge and skills as indicated by their test performance.

We understand that score reports will be developed based on final mockup reports from the Score Report Design Study under the Technology Bundle contract. Pearson will collaborate with the Partnership to produce final report designs for PARCC, state, district, school level reports, school rosters, and student level reports, as described previously in this section under Finalize Design of Reports.

We recognize that, in general, the interpretation guides need to be highly consistent across states, presenting the reports and the interpretations in the same manner. However, given the number and diversity of member states, we anticipate and are fully prepared to accommodate some necessary level of customization.

Such changes might range from relatively superficial differences, for example, in the cover of the guide for different states, to more significant differences such as the specific meaning and interpretation of growth scores from state to state. Any such differences will be carefully instituted and tracked, and an important aspect of our quality control process will involve checking that an interpretation guide customized for a specific state is made available only to that state.

Pearson will develop the interpretation guides in collaboration with PARCC staff on a schedule that provides extensive opportunities for review and discussion of draft materials among appropriate Partnership representatives and cycles of incorporation of resulting feedback into subsequent revision and further review as required. Similarly, feedback from relevant stakeholders (e.g., parents, teachers, and others) will also be solicited as a significant component of initial development such that feedback related to all aspects of the guides as they will function in an operational environment—including accessibility, ease of use, and comprehensibility—will be provided directly from a representative group of potential consumers of reports and interpretation guides from across member states. We will collect this user feedback in multiple cycles and on a timeline that permits consideration and integration of such feedback to inform ongoing revision and enhancement of the interpretation guides prior to their operational implementation.

Interpretive Guides from an Experienced Provider

Pearson will provide interpretive guides for the PARCC assessments that draw on our experience in providing similar documents for a variety of district, state, national and consortium-level clients. We will provide an interpretive guide for parents and educators that will be written at a level that provides these users a comprehensible picture of student capabilities within each area assessed relative to the appropriate educational standards and expectations. This focus will be particularly crucial given the multi-state nature of the PARCC Partnership, and the need to be able to translate the guide into 10 languages to accommodate the rich diversity of populations within member states.

The interpretive guides will be a collaborative effort involving several areas within Pearson, including content, psychometric, program management, editorial, and Creative Services Group. Pearson psychometric staff are recognized for their ability to accurately communicate complex information in clear language for parents, teachers, and other interested stakeholders. We will apply these skills to provide information about cut-scores, proficiency levels and descriptors, scale scores, growth scores, and other report features as appropriate, including how they work, why they are important, and how they can be properly interpreted and understood. Even more importantly, we will help users of this information understand how this information can be used in conjunction with score report data to inform teaching, learning, and assessment practices.

To help the interpretive guides stand out while presenting score information in a clear and uncluttered manner, we will call on our award-winning Creative Services Group (CSG) for creative consulting, concept development, graphic design and illustration, project coordination, and digital color printing processes. Our designers offer innovative solutions for print and digital media projects that other clients have found attractive and easy to read.

To deliver top-notch design to meet the Partnership's publication needs, we use an extensive array of software and digital technology, including today's leading design and publishing tools.

Schedule and Plan for Developing and Implementing Interpretation Guides

Pearson envisions the following steps involved to develop and implement both print and online interactive interpretation guides:

- Pearson coordinates work internally to prepare initial drafts of the proposed interpretive guides, working from the finalized report designs
- Pearson provides initial rough drafts to a working group of appropriate Partnership staff
- Partnership reviews rough draft materials
- Pearson and PARCC working group discuss initial draft materials
- Pearson revises or re-creates guides to accord with feedback from initial review and discussion
- Pearson distributes Version 2 of the proposed guides to PARCC
- PARCC reviews Version 2 and provides feedback
- Pearson makes final adjustments to the draft and re-submits to PARCC
- PARCC provides preliminary approval of draft versions
- Pearson works with PARCC staff to identify an appropriate, representative set of stakeholders (e.g., educators, parents) to review the print-ready and web-based interpretation guides, as well as mock ups of the score reports
- Pearson convene remote focus groups of stakeholders (e.g., via WebEx) to collect feedback on the interpretation guides
- Pearson revises guides based on feedback from the focus groups as needed
- Pearson distributes the final draft version of the proposed guides to PARCC
- PARCC reviews and approves the final draft version (or requires additional edits from Pearson to finalize, as appropriate)

After we have a final set of reports and guides, Pearson will also develop materials for a one-hour webinar to train parents and educators on contents and use of the interpretation guides to view and understand the score reports appropriately. We may collaborate with members of the end-user group to take advantage of their familiarity and experience to help us identify specific focal areas in the training as well as approaches that may be most effective in describing and clarifying areas of the reports and guides.

Once the training webinar has been developed in initial form, we may, if PARCC prefers, offer a pilot training opportunity for a second user group of teachers and parents who have not been involved in the development process. Such a training session would provide an opportunity to evaluate the effectiveness of the materials and the proposed training webinar and to make final adjustments if needed.

Once the webinar materials and process have been developed and finalized, Pearson will conduct a webinar with a new group of end users. This webinar will be recorded and placed on one or more appropriate websites for use throughout the life of the assessment.

Pearson will monitor the interpretation guides throughout each year and provide annual updates to all print and web-based materials to reflect changes needed to them.

V.D.4. Reporting System Training

Requirement

Response Requirements for Section V.D.4.

a) Detailed plans about the objectives and sections of the training.

Deliverables for Section V.D.4.

a) Materials for online training module.
b) Evaluation report of the effectiveness of system training.
c) Annual updates to all systems training materials reflecting any changes to the assessment reports.

Response

Pearson will manage and configure the data management and reporting system developed by the contractor for Component 2 of the PARCC Technology Bundle RFP. As part of the management and configuration, we will develop training resources and offer support services to allow staff from local and state education agencies in PARCC states to effectively use the new system.

The new data management and reporting system will offer sophisticated dynamic reporting capabilities that will require a higher level of training compared to that of current reporting systems used for state assessments. To fulfill the higher training requirements, we propose a solution that utilizes both online modules and live and recorded web-based meeting sessions to provide multi-media, media-rich instruction for both state-level and local users that can be accessed anytime. Adopting a virtual training approach for local users will reduce time and expense spent on training.

Pearson has extensive experience delivering training to educators using numerous online products such as WebEx, Adobe Flash, Adobe Captivate, and Articulate Presenter. We develop training for a variety of media, including self-paced interactive online modules, videotapes, CDs, and DVDs, as well as traditional “stand-up” or in-person training. We have demonstrated capability to provide professional development and training to large audiences, with successful implementations for as many as 180,000 users. Our strategy for training users at a range of technology levels is to develop modular components – small bites that are “digestible” by low-tech users, yet still providing an interactive multimedia training experience.

Online Training Module

Online training will be posted either on the Partnership website, Partnership Resource Center, or other website designated by PARCC. The module will be available in advance of the first reporting window in the fall of 2015.

We will provide one (1) hour of system training to be divided into smaller modules covering step-by-step instructions for using the reporting system. We will also provide an additional 30 minutes of report training covering topics such as how to read and interpret the reports. The modules will be updated as needed to reflect any changes that are made to the data management and reporting system after the initial training modules are released.

As an example of an online training that Pearson has provided for other programs, below is a screen shot of the landing page for the Technology Readiness Tool (TRT) online training module. The TRT is currently used by PARCC and Smarter Balanced schools to capture key indicators for online testing readiness including number and type of computers, external bandwidth and local network infrastructure, and local staff resources.

TechReadinessTool

Welcome to training on the Technology Readiness Tool. This training shows users in all roles how to enter and maintain data in the tool and how to use the reporting feature.

Launch Online Training (with menu)

Click on the links below to view the training with menu navigation. Navigation controls at the bottom of the window allow you to pause or play and to move forward and back within the section you are viewing by clicking and dragging the slider control. Click the arrows or menu items to jump to the beginning of other sections.

Click the icon in the lower right to hide the menu. Click it again to hide most of the navigation controls and enlarge the image. Click the icon a third time to see the menu and controls again.

Training Title	Click to Play
Technology Readiness Tool	
Module 1: Devices, Surveys, and Completion Status - for all users	
Module 2: Organizations and User Accounts - for SRCs and District Admin users	
Module 3: Results and Indicators - for all users	

Click here if you prefer to view the training in the .mp4 file format.

More Information

For more information about the Technology Readiness Tool, review the FAQs in the Answer Center under Support Information at www.TechReadiness.org or contact your State Readiness Coordinator.

Training. Drawing on experience from similar training requirements, Pearson will develop an easy-to-access, complete online training module for the Partnership's online data management and reporting system.

Live WebEx Q & A Sessions

In addition to the online training module, prior to each reporting window, Pearson will conduct up to four web-based training sessions, using Cisco's WebEx product, targeted to state education agency users. Participants will be asked to complete the online training module and submit questions prior to attending the WebEx training. The WebEx sessions will provide an opportunity to review the content in the online module and address any questions, or special state situations or scenarios. The expectation is the state education agency users will waterfall the information from the WebEx sessions to the local levels. Training will be scheduled at various times of day on several days of the week to allow states in all time zones with varying schedules to participate. Key questions action items from those calls will be captured and posted to the training website.

The development of the online training module, content for WebEx sessions, and other training materials will include approval cycles by the Partnership as required.

Requirement

Response Requirements for Section V.D.5.

a) Detailed workflow and data handling plans that depict how the Contractor will integrate software and data management and ensure quality and security controls across PARCC vendors over the life of the contract.

b) Detailed plans for how the Contractor will ensure compliance with federal laws and PARCC policies for the highest level industry standards for data and report privacy and security, including plans on training Contractor's staff and subcontractors on security policies and protocols. The Contractor will be subject to security audits at the discretion of PARCC.

Deliverables for Section V.D.5.

a) Within 30 days of contract execution, the Contractor shall provide a detailed plan and timeline for Reporting workflows and critical milestones.

b) On an annual basis, the Contractor shall provide PARCC with updated technical schemas and documentation for all data elements and reporting formats that are developed and/or enhanced by the Contractor during the period of performance.

Response

Pearson will use the PARCC Technology Bundle’s data management and reporting components to fulfill the reporting requirements and activities required beginning with the 2014–15 school year. We will use the final mockups and specifications to configure and create templates for the reporting system. Once the configurations are complete, the validation process will occur to verify the system and reports. The hosting and maintenance for the system will be provided through a separate procurement process.

Integration Requirements

Pearson will be responsible for management and integration efforts related to software and data handoffs between the components provided through other PARCC vendors such as the field test assessment administration vendor, the data management and reporting technology vendor, and the technical operations vendor, as well as any Pearson-supplied components provided through this contract. This will include responsibility for quality control of data in preparation for the first operational year and future administrations associated with this contract. Pearson will also be responsible for implementation management and support of PARCC’s components.

To integrate software and data management and provide quality and security controls across PARCC vendors over the life of the contract, Pearson will begin by learning directly from the various vendors about their systems and how they work. Based on this knowledge and understanding, we will then develop a detailed workflow, within 30 days of contract execution, based on the points of interaction between discrete vendors’ systems, as these are the vital points where data format standards must be in place for interoperability. We will develop data handling plans that take into account the designs and interfaces of the various vendors' systems to check that the required and contracted data formats are being used and that the required quality and security checks are being performed.

Pearson already works with a number of customers for assessments and complies with known laws and standards as well as customer-specific policies. We will also work with PARCC to comply with federal laws and PARCC policies for data and report privacy and security. Pearson will permit security audits by PARCC and at PARCC's discretion as required by the contract. For more details about security, see the Assessment Administration section.

Annually, Pearson will provide PARCC with updated technical schemas and documentation for all data elements and reporting formats that are developed and/or enhanced by Pearson during the contract period.

The details of Pearson's commitment, including detailed workflow and data handling plans, are as follows:

a. Pearson will work with the data management and reporting technology vendor to establish pre-formatted reporting templates with dynamic reporting interfaces and minor customizable options, such as state logos, for the 2014–15 Operational Assessments.

b. Pearson will work with PARCC to inform upgrades and improvements to the data management and reporting system for each subsequent year of the contract.

c. Pearson will use the interfaces and reporting tools built by PARCC's data warehouse and reporting technology vendor to generate new templates, ad hoc reports, and report modifications.

d. Pearson will use the interfaces and reporting tools built by PARCC's data management and reporting technology vendor to generate all necessary reports, administer quality control procedures, ready reports for distribution to states and other authorized entities, and archive reports in the PARCC Data Warehouse in accordance with Partnership data policies.

e. Pearson will adhere to Partnership policies and protocols for data and respect privacy and security, including integration with PARCC's user identity management / single sign-on system that will be used to establish role-based permissions for data access, import/export, and reporting.

f. Pearson will work with Partnership representatives to establish and/or refine the data elements and reporting formats to be included for each different report.

g. Pearson will work with PARCC's data management and reporting technology vendor to achieve the successful development and deployment of dynamic and static reports in accordance with Partnership policies and protocols for data and report privacy and security.

Interoperability Requirements

All PARCC reports generated through Pearson activities will adhere to interoperability guidelines including schemata being established by the Partnership using the Assessment Interoperability Framework (AIF), Common Education Data Standards (CEDS), and School Interoperability Framework (SIF) standards. Pearson will work with Partnership representatives to identify necessary data elements for different reports, and will obtain PARCC review and written approval of data model and data transport formats to be used as a part of Operational Assessment administration activities and deliverables.

V.D.6. Paper-Based Student Reports

Requirement

Response Requirements for Section V.D.6.

a) Plans for student report delivery

Deliverables for Section V.D.6.

a) Within 30 days of contract execution, the Contractor shall provide a detailed plan and timeline for delivering printed student reports and critical milestones.

b) Printed and delivered paper-based student reports.

Response

Clear, Concise Reports to Communicate Student Performance Information

The main purpose of the reports produced for PARCC assessments is to enhance learning at the student level and to provide information about proficiency at the classroom, school, and district levels. Therefore, it is not enough to just generate reports. They must be comprehensible, reflecting student performance in a clear and concise fashion. We will design reports to maximize their usefulness at all levels.

As part of the reporting requirement process described in Section V.D.1 we will work with PARCC to identify the need for printed reports for each operational year.

Following final approval of report requirements and system development, we will perform a system test of report production, just as we do a system test before full production scanning. In the system test of report production, the Pearson production planning analyst works with the Pearson PARCC Program Team and with the Printing and Output Processing (reports assembly) areas in Operations to identify a test production run of report printing and assembly. After we validate the reports against PARCC requirements, we will send sample hard-copy reports for PARCC review and approval. We will provide a sufficient number of working days for Partnership staff to review and approve the reports.

Shipped Promptly to Districts/Schools

Trained shipping personnel will determine the most reliable and rapid means of delivering each shipment of reports. We will enter each district's/school's reports in the shipping manifest system as they are shipped. The shipping manifest system allows shipments to be traced quickly should delivery issues arise.

Pearson will select the mode of transport that best fits the district/school location, quantity of the shipment, and amount of time the shipment has to reach the district our Cost Proposal for this state option is based on Ground Transportation.

V.D.7. Quality Control

Requirement

Response Requirements for Section V.D.7.

a) Quality control specifications and related materials that describe in detail all of the steps to be implemented to demonstrate to the Partnership that the reports are accurate.

Deliverables for Section V.D.7.

a) Regular updates on quality control processes.

Response

Commitment to Accurate Reporting Results

To measure the quality control performance of our operations units we use consistent business-wide metrics to evaluate and communicate our success in meeting our delivery objectives. The key performance metrics tracked for reports printing and assembly include the following:

- **Report Quality.** Measured by defects per million
- **On-Time Delivery.** Measured by percent delivered on time with a performance target of 100 percent on-time delivery.

In 2012, Pearson provided reports for 332 projects (administrations). Ninety-six of those projects were shipped on the planned shipment date. Even when the planned ship date wasn't met, all 332 projects received their reports within the planned reports delivery timeframe.

Testing for Print Quality

Pearson's Printing Operations follows a first production run test process to verify the high quality of its printing jobs. The first production run is conducted to confirm that a job will run smoothly once it is put into regular production. Every printing project that requires printing for results reporting must be approved for production following this process. The first production run process (blue dot) is required for both printed and non-printed (media) output generated by Printing Operations.

To confirm the print job performs each function as expected, printing operators perform the functions listed on a standard checklist. If the test job is unable to perform an item on the checklist, the operator indicates the reason for the failure. The list is then returned to the appropriate production control analyst for problem resolution before the job is released into full production. Sign-off and acceptance by the program team, software and technology services, and Operations printing operators is required prior to full report production.

Reports processing involves assembling and distributing students' test results in a report and/or media format to states, districts, and schools throughout the country.

- Individuals assembling will folder reports based on project specifications and packing lists.
- The assembler will also do a print quality check as the reports are being foldered to verify the reports are readable.
- Depending on the project, assemblers may need to verify that certain data is present on reports.
- After the reports are assembled, a quality control check is completed, verifying all reports are present and have been packaged and labeled correctly.
- All issues are documented and feedback is provided to the assembler responsible for the issue. Whenever possible, the assembler will correct his/her work. Data is maintained of all types of issues and of the individuals who were responsible. This information is used for determining trends and areas needing improvement, to further enhance training, and to implement processes or changes.

Shipping Quality Checks

All materials being shipped will go through the Output Processing full service shipping area.

- The individuals boxing will verify that the mailing label matches the materials being shipped. When provided, the boxer will also compare the label to the OSS.
- The shipping clerk manifesting the boxes will scan the mailing label, which will generate a shipping label. This label is then QC'd against the mailing label to confirm the correct shipping information has been produced.
- If any shipping issue occurs, the issue is documented and feedback is provided to the shipper or individual responsible and immediate steps are taken to correct the issue.
- Data is maintained of all issue types and of individuals responsible. This information is used for determining trends and areas needing improvement, to further enhance training, and to implement processes or changes.

Delivering Reports Accurately. For accurate processing and shipment of reports, Pearson uses the following **production validation** steps to verify the packaging and shipment of live materials:

- As with test materials packaging and distribution, our Organizational Quality group performs **pre-blue dot** verification to confirm that pre-production activities are complete before the blue dot process and to confirm the operations departments' preparation for generating the production materials. This process entails generating all reporting deliverables in the production environment.
- In the **blue dot process,** based on specific demographic criteria, we select a sample of districts awaiting reports. We produce reports in the production environment as if they would be sent to the districts. After packaging pre-blue dot reports, the Quality Control staff verifies them for accuracy, completeness, and adherence to PARCC requirements.
- **We randomly sample reports** during assembly and packaging to confirm adherence to specifications throughout the distribution.

The following figure illustrates our quality assurance process for reports.

Quality Reporting Process

1. **Unit and Production Configuration Testing**
   - Software Developer and the IT Assessment Testing group perform extensive unit testing on reporting software programs
   - Software Developer and the IT Assessment Testing group perform extensive Product Configuration Testing on reporting software programs

2. **Acceptance Testing**
   - Performed by the independent Quality Group (QG)
   - Process the same scored data through the reporting software
   - All output is verified noting that products are integrated properly
   - Report deliverables are verified that they meet customer expectations.

3. **Production Validation**

   Pre-Blue Dot
   - Performed by QG to verify all pre-production activities have been completed.
   - QG verifies that all operation departments are prepared to generate the production materials.

   Blue Dot
   - Specific districts containing pre-determined criteria needed to verify proper report generation are processed.
   - QG verifies all reporting deliverables for accuracy, completeness and print quality.
   - Verify reporting deliverables meet customer expectation

   Report Packaging Spot-Check
   - Random sampling during packaging to verify adherence to specifications

Requirements-Driven Quality Control. Detailed PARCC reporting requirements serve as the backbone of our end-to-end quality control process.

Data Uploads, Reporting and System Quality Control

Pearson will monitor all aspects of the reporting procedures, including key-based and rule-based machine scoring and handscoring for constructed-response items and performance tasks. Pearson has tools and processes in place to monitor procedures throughout the data preparation, integration, scoring, analysis, and reporting processes.

The Pearson validation team will prepare test plans used throughout the scoring and reporting process. These test plans will include details for, at a minimum, loading scored data into the reporting system, reporting data and formats, and general system maintenance. Test plan preparation is typically organized around detailed specifications for the following areas:

- Raw score validation (e.g., score key validation, objective/strand/domain scoring, field test non-score, double-grid combinations, possible correct combination, if applicable, and out-of-range/negative test cases)
- Derived scoring, if applicable (e.g., scaled score, performance level, and percentile score validation)
- Matching (e.g., validation of high-confidence criteria, low-confidence criteria, cross-document, external or forced matching by customer; prior to and after data updates; extract file of matched and unmatched documents)
- Demographic update tests (e.g., verification of data extract against corresponding layout, valued values for updatable fields, invalid values for updatable/non-updatable fields; negative test for non-existing record or empty file)
- Aggregation, if applicable (e.g., tests of summary report data and field-level calculations, inclusion and exclusion criteria, minimum and maximum values for reporting categories, population subset confirmation, effects of attemptedness on aggregation outcomes)

This is a standard part of the ongoing quality control processes we engage in throughout our program execution activities. We have well-developed procedures and policies in place to promote thorough examination of our processing throughout these stages, and identification and resolution of any issues that might arise.

In addition, the mock data and processes described in the Assessment Administration section test that our systems are operating according to agreed-upon specifications.

Pearson will provide PARCC with regular updates on quality control processes.

Component 5: Standard Setting

V.E.1. Overview and Timeline

Requirement

V.E.1.A. Performance Level Descriptors
V.E.1.B. Challenges
V.E.1.C. Preferred Method

Response

Overview of Performance Level Descriptors, Preferred Method, and Challenges

Performance Level Descriptors

Performance Level Descriptors (PLDs) will be used to inform the range of task and item complexity required at each grade level to measure the full range of proficiency as well as to provide maximum discrimination between performance levels. The evidence statements, task generation models and rubrics must all be aligned with and reference the PLDs.

It is expected that there will be multiple iterations of the PLDs. The PLDs must evolve to deeper levels of specificity as more information is known about student performance on PARCC assessment tasks. The current PLDs can be found at http://parcconline.org/plds, and the PLDs will be further refined after the first operational administration when performance standards are set, and after 2-3 years of the operational assessment when additional empirical data are available. ETS will work with Partnership Manager representatives to update PLDs, evidence statements, task generation models and rubrics as necessary, and keeping these documents aligned across grades within a subject area. The PLDs, assessment specifications, and supporting documents are “living” documents and shall be updated, in consultation with the Partnership Manager, as needed or required as additional test results and research and validity data becomes available.

Preferred Standard Setting Method

The standard-setting process for the PARCC summative assessments will integrate the PARCC's College- and Career-Ready (CCR) Determination Policy, policy-level and subject- and grade-specific Performance Level Descriptors (PLDs), empirical data, and content expert judgment to set five performance levels for each assessment.

The CCR Determination Policy describes students' readiness for college and careers for the courses in grade 11. The performance levels for grades 3–11 discuss students' readiness for future grades or in the case of grade 11, students' readiness for college and careers. The use of empirical studies in the standard setting process to support the policy statements aligns well with PARCC's inclusion of student readiness in the definition of the performance levels.

Pearson proposes an Evidence Based Standard Setting process (EBSS; McClarty, Way, Porter, Beimers & Miles, 2013) to integrate empirical data from systematic research and content expert judgment in setting the performance standards for PARCC. EBSS is a process that supports policy claims through systematic research designed to inform the judgments made by content experts. EBSS lends itself well to creating a system of aligned performances standards starting with college and career readiness and linking down from high school to middle school to elementary school.

This approach has been used successfully to set performance standards on several assessment programs in recent years, including the American Diploma Project (ADP) and the New York and Texas assessment programs.

The critical elements of the proposed EBSS approach are described below.

- **Curriculum.** The CCSS, which provide a clear and consistent description of the knowledge and skills students are expected to learn, serve as the underlying basis for several key components of the standard-setting process, including the definitions for each performance level and grade/content area-specific PLDs.
- **Assessment.** Each PARCC assessment has been developed to assess the knowledge and skills described in the CCSS. Each PARCC assessment is based on the requirements described in the anchor standards for the specific grade level and content area and should adhere to the published blueprint and test specifications.
- **Policy Considerations and External Validation.** Results from research studies, which (1) compare performance on the PARCC assessments with scores on other related measures or external assessments and gather post-secondary instructors' judgments on CCR expectations. Stakeholders and experts from across the member states with experience in educational policy and knowledge of the PARCC assessments consider the study results when making recommendations about reasonable ranges for setting each performance standard.

- **Expertise and Knowledge about Students and Subject Matter.** Educators, including classroom teachers and curriculum specialists from K–12 and higher education across the member states, bring content knowledge and classroom experience to the standard-setting process. They play an integral role in developing the PLDs and in recommending the performance standards.
- **Standard Setting.** Within the framework of EBSS, an established standard-setting method known as the bookmark method with external data (Ferrara, Lewis, Mercado, D'Brot, Barth, & Egan, 2011; Phillips, 2012) is used to recommend the cut scores for each PARCC assessment.

To implement the EBSS approach for the PARCC assessments, we propose the following seven-step process.

1. Define the outcomes of interest and policy goals
2. Develop research, data collection, and analysis plans
3. Synthesize the research results
4. Conduct pre-policy meeting
5. Conduct standard-setting meeting with panelists
6. Conduct reasonableness review through post-policy meeting
7. Continue to gather evidence in support of the standards

PARCC is positioned well for the use of an EBSS process through the definitions of the CCR Determination Policy, the policy level PLDs, and subject- and grade-specific PLDs. The EBSS approach best fits the needs of PARCC because, if properly planned and executed, it will result in performance standards that not only represent the students' degree of mastery of the assessed curricula based on the CCSS, but can also be used to evaluate their readiness for college and careers.

Challenges

Based on Pearson's experience with implementing the EBSS approach in support of the American Diploma Project and in Texas, we understand that the challenges associated with this work are most likely to be related to the timing and logistics of working across many states to gather student-level data for the research studies, and recruiting representative panels for the PLD meetings, research studies, and standard setting events. Additionally, each state and/or institution is likely to have distinct privacy concerns that will require different pathways for collecting information. There is also likely to be different governance concerning participation incentives at the institutions selected to participate in the research studies in support of the standard setting events. Each of these challenges is understood by Pearson and we are prepared to successfully overcome them.

V.E.2. Performance Level Descriptors for Standard Setting

Requirement

Response Requirements for Section V.E.2.

a) Description of the approach and procedures to complete all the responsibilities/tasks specified in Section V.E.2

b) Description of approach in revising and synthesizing current PLDs to make them ready for use in standard setting

c) Description of approach in ensuring that the revised PLDs are vertically moderated across all grades within each content area.

Deliverables for Section V.E.2.

a) Plan to revise and synthesize standard setting PLDs to be approved by PARCC

b) Draft PLDs revised and synthesized as described above, and other materials for six four-hour webinars to be conducted with the original PLD panelists for purposes described above by 6/20/2014

c) Final standard setting PLDs revised and synthesized as described above by 8/4/2014

Response

Preparing Performance Level Descriptors for Standard Setting

At the heart of PARCC's standard setting activities are the performance level descriptors that will be used for PARCC College and Career-Ready determinations. PARCC's policy-level performance level descriptors (PLDs), which were approved by the PARCC Governing Board in October 2012 and the grade- and subject-specific PLDs, which were approved by the PARCC Governing in June 2013), represent a critical part of the chain of validity evidence used in building and defending the PARCC assessments.

It is recognized in standard-setting research and practice that PLDs form the foundation for standard setting panelists' judgments (e.g., Egan, Ferrara, Schneider, & Barton, 2009; Perie, 2008). We are aware that PARCC developed subject- and grade-specific PLDs through an iterative process that included higher education and classroom teachers with relevant content expertise and that this iterative process will continue as data become available upon completion of the item tryout, field test, and first operational administration. ETS was very pleased to be able to provide content and measurement expertise to support to PARCC in the development of its College and Career-Ready performance level descriptors. This relationship has provided the opportunity to maintain alignment among the critical pieces of assessment development such as task generation models, rubrics and test claims with expectations stated in the PARCC PLDs.

In preparation for PARCC's standard setting activities, we propose to leverage ETS' experience with the development of PARCC's PLDs by leading meetings with PLD panelists to produce content and grade level PLDS for use in standard setting. The PLD revision process will include the panelists involved in the original PLD development process prior to the standard-setting panel meetings. We plan to convene seven grade-band PLD panels: three ELA and four mathematics panels. The panelists will meet six times via webinar, each of which will be four hours in duration.

During the first session, ETS will provide training to the panelists to orient them to the process and their assigned tasks and facilitate subsequent sessions with panelists. An outline of the major activities that will follow the initial training session is outlined below.

- Grade-band panels of seven members will first work within each grade in the grade-band to establish reasonable expectations within-grade across level and will proceed to verify articulation across grades
- PLD panel facilitators will provide feedback from the within-grade-band meetings to panelists during subsequent cross-grade-band meetings to allow members to review results across grade bands. Reviews of cross-grade-band PLDs will occur so that each builds upon the earlier work.
- Three cross-grade panels will meet to discuss adjacent grades initially (grades 3–5, 6–8, and high school) and modify expectations as needed to develop vertically articulated descriptors within these groups. A subsequent review of the bridging grades (grades 5–6, and grade 8 with high school descriptors) will occur to verify vertical articulation across the full grade span.
- ETS assessment development content experts will synthesize the feedback from the PLD panels after each webinar, working with the panels in an iterative fashion
- Each draft of PLD documents will incorporate the discussion by the panel, and each review by the panel will further validate the PLDs.
- The result of this process will be PLDs that indicate increasing expectations across performance levels within each grades, are vertically articulated across grades, and are streamlined and focused toward cut scores placed on the overall score scale

The PLDs will be available by June 20, 2014 for the PARCC standard-setting meetings. As noted, the integrity of PARCC's standard setting PLDs is critical to the quality of the standard-setting process and ultimately to the validity of PARCC scores. ETS understands that the development of PARCC PLDs must be an ongoing effort, aligned with assessment development activities. PLDs are drivers in the test development process; test claims are considered throughout test design and task development activities, and the alignment of PLDs with the development of evidence statements, task generation models, and scoring rubrics must be checked with vigilance. The current coordination between PARCC and the item development vendor is a good model for maintaining this alignment, because it provides a close link between the assessment components and the expectations provided in the Performance Level Descriptors.

After the operational assessment has been delivered and student data are available, further refinements to the PLDs may be required. The assessment development content experts will be most knowledgeable about the test content and ideally suited to make meaningful modifications and maintain alignment.

V.E.3. Logistical Requirements

Requirement

Response Requirements for Section V.E.3.

a) Description of the approach and procedures to complete all the responsibilities/tasks specified in Section V.E.3

b) Description of staff experience

c) Description of approach in ensuring that the revised PLDs are vertically moderated across all grades within each content area.

Deliverables for Section V.E.3.

a) Timeline for arranging logistical details of workshop

b) Detailed workplan for all activities that lead to the approval of final standard setting method(s) by the Partnership

Requirement

V.E.3.A. Standard Setting Workshop

Response

Meeting Planning for Standard Setting Workshops

Standard Setting Workshops will be held following the first operational year of the PARCC Assessments. The meetings will be held in a hotel, in a PARCC state, near a major airport hub. For cost planning purposes we have assumed that the workshops will be held near Chicago O'Hare Airport or another city in a PARCC state with a similar GSA rate. Given the size of the meetings—154 participants for the largest workshop, including contractor staff—we will select a hotel that has demonstrated capability to handle a large number of simultaneous computer and internet users.

Below are key assumptions about the workshops, taken from the RFP and Amendment 1. We have also included the number of staff from Pearson, ETS, and WestEd that will attend the workshops, in the last column of the table. Based on our experience, this level of staffing will be required, and exceeds the number requested by PARCC on page 174 of the RFP.

Each workshop will be broken up into multiple panels by content area and grade/level. Pearson will reserve a ballroom for training workshop participants, with separate breakout rooms also provided.

Key Assumptions for Standard Setting Workshops							
Meeting	Dates	Meeting Duration (Days)	# of State Participants	# of PARCC Staff Attending	Total Travelers (State + PARCC)	Number Participants Eligible for Stipends/Substitutes per Meeting	Number of Contractor Staff Attendees
Standard Setting For High School Assessments	7/27/15-7/31/15	4	123	8	131	123	23
Standard Setting for grades 7-8	8/17/15-8/21/15	4	43	8	51	43	11
Standard Setting for grades 3-6	8/24/15-8/28/15	4	83	8	91	83	20

Assumptions for Standard Setting Workshops. Our experience in conducting workshops allows Pearson to project how many attendees will be required to achieve PARCC goals.

Whereas many of the PARCC meetings occur in each year of the contract, standard setting is only planned following the first operational year. We have updated the "Travel and Meetings" tab to reflect this assumption.

We share PARCC's expectations for maintaining item/task/text security.

- Participants will be required to sign the PARCC test security and confidentiality (non-disclosure) form, if they have not already done so within the previous twelve months. Pearson will store the forms for up to three years, unless instructed otherwise by PARCC.
- Participants may view items/tasks/text online during the meeting. Pearson will provide a firewalled network for the purposes of the standard setting. In addition, any items viewed online will be through a secure, password protected environment.
- Any hard copies of test items, tasks, and reading passages being reviewed, will be provided to each participant in a security-controlled notebook. The materials will be secured, checked-out when in use and checked back in during use. Secure materials will not be permitted to leave the meeting room, cannot be photocopied, and will be stored in a locked location before and after meetings. Materials will be shipped from Pearson to and from the meeting site using overnight shipping with tracking capabilities. Any materials that can be discarded after a meeting will be shredded.

Providing Necessary Reference Materials

For the in-person meetings, each participant will be issued a laptop to access the necessary information. Hard copies of the following reference materials will also be available for participants to share: hard copies of PARCC Assessments and answer documents, copies of the Common Core State Standards, test item specifications, textbooks, and other materials identified by Pearson and PARCC during the Standard Setting Planning Meetings. Supporting materials such as audio/visual equipment, meeting supplies such as flip charts, note pads, pens, pencils, and other office supplies will also be provided.

Providing Stipends and Substitute Reimbursement

As indicated on page 69 of the RFP, “as a general rule, stipends ($150 per meeting day) shall be paid for meetings that occur in July and/or August.” This amount will be provided to the participants that are eligible for a stipend, with the exception of the three TAC members that are expected to attend each Standard Setting Workshop. As required in the RFP, Pearson will pay TAC members at the rate of $1,500 per meeting day.

We will be responsible for airfare, rental car (if needed), lodging, and meeting meals (breakfast and lunch). Lodging and airfare will be booked directly through Pearson. Participants will be provided with a per diem to cover the expense of dinner, transportation to and from the airport (if needed), mileage, tolls, and other travel related expenses. Per Diem rates will be based on the GSA per diem rates for the city in which the meeting is held. For costing purposes, our proposal assumes Chicago GSA per diem rates.

Pre-Meeting Requirements

As described in Section V.A.1.G, “we will provide PARCC with a Meeting Planning Document at least eight weeks in advance of the first Standard Setting Workshop that provides the following information for each workshop.”

- Meeting location (city and hotel name)
- Date and time
- Meeting agenda
- Participant names, including meeting facilitators
- Requirements for training materials
- Requirements for reference materials and supplies
- Instructions for participants to book their travel
- Nearby dining options

Given the large size of the meetings and the critical need to have the meetings so that cut scores can be set later in the fall, it will be important to start recruiting participants for the meeting more than six weeks prior to each workshop (six weeks was the requirement in the RFP for inviting participants for othe rmeetings). Early in the spring of 2015, we will work with PARCC and State Departments of Education to begin identifying participants. There are three Standard Setting Planning Meetings that will be held in the first year and one in the second year. We will use information discussed during these meetings to develop daily meeting requirements including:

- Setup time
- Meeting times
- Meeting room name
- Room setup requirements (classroom, u-shaped, etc.)
- Audio/visual requirements
- Food/beverage requirements

Section V.A.1.G includes additional information about project specification forms that we will use to capture and communicate requirements regarding the Standard Setting Workshops.

Timeline for Arranging Logistical Details for Workshops

Below is a draft schedule for arranging the logistical details for the Standard Setting Workshops. These timeframes/dates are also included in our Draft Project Schedule in the Other Materials section of this proposal. The Standard Setting Planning Meetings are in-person meetings at a Pearson location and include participants from PARCC, participating states, and Pearson. Depending on the agenda topics for a particular meeting, participants from WestEd and ETS may also join the meeting to discuss their role in meeting facilitation.

Standard Setting Workshops	
Milestone	**Date/Timeframe**
Standard Setting Planning Meeting	April 2014
PARCC Standard Setting OWG and TAC Approval	May 2014
Governing Board/ACCR review/approve final standard setting method	June 2014
Standard Setting Planning Meeting	Summer 2014
Standard Setting Planning Meeting	Spring 2015
Standard Setting Planning Meeting	Summer 2015

Schedule for Standard Setting Workshops. Holding these in person meetings at a Pearson location allows us to control costs while providing an enhanced level of security.

Requirement

V.E.3.B. Staff Experience

Response

Description of Staff Experience

Standard Setting Lead

Pearson proposes Dr. Julie Miles to lead the standard setting. Dr. Miles has more than 10 years of experience leading and facilitating standard settings for large-scale state testing projects. As summarized in the following figure, to date Dr. Miles has designed and led 16 different standard setting events, and she has facilitated in a total 40 different standard setting events. Her experience spans all grades, all subjects, and assessment systems requiring regular, modified, and alternate standards.

In 2009, Dr. Miles designed and led the standard setting for the American Diploma Project Algebra II exam, which involved an innovative, data-based standard setting approach (see http://researchnetwork.pearson.com/wp-content/uploads/AmericanDiplomaProjectAlgebraII.pdf.) This work has led to several peer-reviewed publications in recent years (Haertel, Beimers & Miles, 2012; O'Malley, Keng & Miles, 2012; McClarty, McClarty, Way, Porter, Beimers & Miles, 2013).

Standard Setting Meetings Designed and/or Facilitated by Dr. Julie Miles			
Type	**Year**	**State**	**Program**
Regular	2004	AR	Facilitator for Arkansas Comprehensive Testing, Assessment, and Accountability Program grades 3–8 (Questar, lead)
ESL	2004	ELDA	Facilitator for English Language Development Assessment (Dr. Mike Bunch lead)
ESL	2005	ELDA	Facilitator for English Language Development Assessment (Dr. Mike Bunch lead)
Grad	2005	OH	Facilitator for Ohio Graduation Tests (OGT) (Dr. Elliot Inman lead)
EOC	2006	NC	Facilitator for North Carolina End-of-Grade mathematics tests grades 3–8 (Dr. Paul Nichols lead)
Mod	2006	VA	Lead facilitator for Virginia Standards of Learning, Modified Diploma
Alt	2006	VA	Lead facilitator for Virginia Alternate Assessment Program, reading/math/science grades 3–8, HS
Alt	2006	VA	Lead facilitator for Virginia Grade-Level Alternative, reading and mathematics, grades 3–8
EOC	2007	NC	Facilitator for North Carolina End-of-Course Tests, algebra I, algebra II, geometry, and English I (Dr. Kelly Burling lead)
Alt	2007	VA	Lead facilitator for Virginia Alternate Assessment Program,

Standard Setting Meetings Designed and/or Facilitated by Dr. Julie Miles			
Type	**Year**	**State**	**Program**
			reading/math/science grades 3–8, HS
Mod	2008	MD	Facilitator for Maryland Modified high school assessment, algebra I, biology and English (Dr. Ian Little lead)
Alt	2008	MD	Facilitator for Alternate Maryland School Assessment, reading/math grades 3–8, 10 and science 3, 5, 8 and 10 (Dr. Ian Little lead)
Alt	2008	NC	Lead facilitator for North Carolina NCExtend2 OCS writing grade 10
Regular	2008	VA	Lead facilitator for Virginia Standards of Learning, science grade 8.
Alt	2008	VA	Lead facilitator for Virginia Grade-Level Alternative, science grades 3, 5, 8
Regular	Feb 2009	ADP	Lead facilitator for American Diploma Project Algebra II End-of-Course Exam (regional judgment event #1)
Regular	Mar 2009	ADP	Lead facilitator for American Diploma Project Algebra II End-of-Course Exam (regional judgment event #2)
Regular	Apr 2009	ADP	Lead facilitator for American Diploma Project Algebra II End-of-Course Exam (regional judgment event #3)
Regular	Jun 2009	ADP	Lead facilitator for the American Diploma Project Algebra I and Algebra II End-of-Course Exam (standard setting event)
Grad	2009	MN	Lead facilitator for the Minnesota GRAD, mathematics
Alt	2009	NJ	Facilitator for the New Jersey APA, reading/math/science 3–8 (Dr. Paul Nichols lead)
Regular	2009	PR	Facilitator/Data Analyst for the Puerto Rico PPEA (Dr. Steve Fitzpatrick, lead)
Alt	2009	PR	Facilitator/Data Analyst for the Puerto Rico PPAA (Dr. Steve Fitzpatrick, lead)
Regular	2010	TN	Facilitator for the Tennessee ACH, reading/math/science/history, grades 3–8 (Dr. Erika Hall lead)
Alt	2010	TN	Facilitator for the Tennessee MAAS, reading/math/science/history, grades 3–8 (Dr. Erika Hall lead)
EOC	2010	VA	Lead facilitator for the Virginia SOL EOC history tests
Grad	2011	NY	Facilitator for the NY Regents Exam (Dr. Ye Tong lead)
Regular	2011	VA	Lead facilitator for the Virginia SOL grade 3 history and content-specific history tests
Alt	2011	VA	Lead facilitator for the Virginia VGLA grade 3 history and content-specific history tests
Alt	2011	VA	Lead facilitator for the Virginia VAAP history and writing tests
EOC	2011	VA	Lead facilitator for the Virginia SOL EOC algebra I, algebra II, and geometry and vertical articulation
Reg/Alt	2011	MN	Facilitator for the Minnesota MCA-II and MTAS (Dr. Ye Tong lead); Lead facilitator for vertical articulation of mathematics
Regular	2011	FL	Floating facilitator for Florida FCAT 2.0 reading grades 3–8 (Dr. Rob Kirkpatrick lead)
Regular	2012	VA	Oversight, floating facilitator for the Virginia VASOL grades 3–8 mathematics
Alt	2012	VA	Oversight, floating facilitator for the Virginia VMAST grades 3–8 mathematics and algebra I

Standard Setting Meetings Designed and/or Facilitated by Dr. Julie Miles			
Type	**Year**	**State**	**Program**
Alt	2012	VA	Oversight, floating facilitator for the Virginia VAAP grades 3–11 mathematics
EOC	2012	VA	Oversight for the Virginia VASOL grades 11 reading, writing, biology, chemistry, and earth science
Regular	2012	VA	Oversight for the Virginia VASOL grades 3–8 reading, grades 5 and 8 writing, and grades 3, 5, and 8 science
Regular	2013	NY	Facilitator for the New York mathematics grades 6–8 (Dr. Ye Tong lead); Lead facilitator for vertical articulation of mathematics

Pearson Psychometric Staff. Dr. Miles has experience leading standard setting events for all grades, subjects, and assessment systems including regular, modified, and alternate standards.

Dr. Miles comes highly recommended by Dr. Andy Porter (American Diploma Project), Shelley Loving-Ryder (Virginia), and Jennifer Dugan (Minnesota). For details of their involvement with standard settings led by Dr. Miles, see Attachment V.

Program Manager

Pearson proposes Monica Lyons as the program manager for the standard setting event. Ms. Lyons will also serve as the program manager for the associated research studies so that there is continuity across the required tasks leading up to the event.

Group Facilitators

Pearson proposes that each standard setting panel be led by a psychometric facilitator with expertise in standard setting along with support from a content facilitator with intimate knowledge of the PARCC assessments. This psychometric-content team will share the responsibility for facilitating the individual panel discussions and insuring that the standard setting process is followed consistently within and across rooms and events to support procedural validity.

Pearson will provide psychometric leads with extensive standard setting expertise for each panel. Depending on availability at the time of the event, the following figure shows the standard setting experience of our offered staff. Pearson will work with the PARCC Governing Board to verify that the appropriate facilitators are selected.

Proposed Pearson Psychometric Facilitators		
Name	**Title**	**Standard Setting Experience**
Tracey Hembry	Manager, Measurement Services	▪ North Carolina, 2013 End of Grade ELA grades 3–5 ▪ Arizona, 2013 Arizona English Language Learner Assessment (AZELLA) Stage I (Kindergarten) ▪ Virginia, 2013 Standards of Learning writing grade 5

Proposed Pearson Psychometric Facilitators		
Name	**Title**	**Standard Setting Experience**
		▪ Texas, 2012 State of Texas Assessments of Academic Readiness (STAAR) Alternate algebra I and geometry ▪ Texas, 2012 State of Texas Assessments of Academic Readiness (STAAR) Spanish writing grade 4 ▪ Texas, 2012 State of Texas Assessments of Academic Readiness (STAAR)writing grade 7 ▪ Pennsylvania, 2012 Pennsylvania Educator Certification Tests (PECT) ▪ California, 2011 California Preliminary Administrative Credential Examination (CPACE) ▪ Oregon, 2010 National Evaluation Series (national teacher certification tests) Oregon standard setting ▪ National, 2010 National Evaluation Series (national teacher certification tests) National standard setting ▪ Minnesota, 2010 Minnesota Teacher Licensure Examinations (MTLE) ▪ California, 2009 Reading Instruction Competence Assessment (RICA) ▪ Massachusetts, 2009 Massachusetts Tests for Educator Licensure (MTEL)
Leslie Keng	Manager, Measurement Services	▪ Texas, 2007 Texas Assessment of Knowledge and Skills–Alternate (TAKS–Alt) reading, grade 6–8 ▪ Texas, 2007 TAKS–Alt social studies, grade 8, 10, 11 ▪ Mississippi, 2008 Mississippi Curriculum Test, Second Edition (MCT2) Language Arts, grade 8 ▪ Texas, 2008 Texas Assessment of Knowledge and Skills–Modified (TAKS–M) reading/Language Arts, grade 9–11 ▪ Texas, 2008 TAKS–M science, grade 5 and 8 ▪ Texas, 2008 TAKS mathematics, grade 6–8 ▪ Tennessee, 2010 Modified Academic Achievement Standards (MAAS) reading, grade 3–4 ▪ Tennessee, 2010 Achievement Test (ACH) reading, grade 3–4 ▪ Florida, 2011 Florida Comprehensive Assessment Test 2.0 (FCAT) reading, grade 10 ▪ Texas, 2012 State of Texas Assessments of Academic Readiness (STAAR) English I, II, III reading ▪ Texas, 2012 STAAR 3–8 reading, grade 3–5 ▪ Texas, 2012 STAAR 3–8 writing, grade 4 ▪ Texas, 2012 STAAR alternate reading, grade 3–5 and writing, grade 4 ▪ Texas, 2012 STAAR Modified English I and II reading ▪ Texas, 2013 Texas English Language Proficiency Assessment System (TELPAS) reading, grade 4–5, 6–7 ▪ Virginia, 2013 Virginia Alternate Assessment Program (VAAP) science, grade 3, 5, 8 and high school ▪ New York, 2013 New York State Teacher Certification Examinations (NYSTCE), School Building Leader Part 1 (SBL1) and School Building Leader Part 2 (SBL2)

Proposed Pearson Psychometric Facilitators		
Name	**Title**	**Standard Setting Experience**
Katie McClarty	Director, Center for College and Career Success	▪ Texas, 2012 State Assessments of Academic Readiness (STAAR) mathematics, grade 4 ▪ Texas, 2012 STARR mathematics, grade 5 ▪ Arizona, 2011 Arizona Instrument to Measure Standards (AIMS) writing, grade 7 ▪ Texas, 2008 Texas Assessment of Knowledge and Skills (TAKS) Modified English Language Arts, grades 9–11 ▪ Texas, 2008 TAKS standards review Spanish mathematics, grades 3–6 ▪ Texas, 2007 TAKS Alternate science, grades 5 and 8
Sonya Powers	Research Scientist	▪ Virginia, 2011 Virginia Grade Level Alternative (VGLA) Virginia Studies ▪ Arizona, 2011 Arizona Instrument to Measure Standards (AIMS), writing, grade 5 ▪ Texas, 2012 State Assessments of Academic Readiness (STAAR) End-of-Course US history ▪ Missouri, 2012 Missouri Assessment Program Alternate (MAP-A) science, grade 8 ▪ Texas, 2012 STAAR Alternate mathematics, grades 3, 4, and 5 ▪ Texas, 2012 STAAR Spanish reading, grades 3, 4, and 5 ▪ Texas, 2012 STAAR Modified social studies, grade 8 ▪ Texas, 2012 STAAR Spanish writing, grade 4 ▪ Arizona, 2013 Arizona English Language Learner Assessment (AZELLA) high school ▪ Texas, 2013 Texas English Language Proficiency Assessment System (TELPAS) reading, grade clusters 8–9 and 10–12
Natasha Williams	Director, Psychometric and Research Services	▪ Texas, 2013 Texas English Language Proficiency Assessment System (TELPAS) reading 2–3 (including general session for all grades), grades 2–3 ▪ Texas, 2012 State of Texas Assessments of Academic Readiness (STAAR) End-of-Course, algebra I, algebra II, and geometry ▪ Texas, 2012 STAAR Modified General Session for all grades/subjects ▪ Texas, 2012 Alternate General Session for all grades/subjects ▪ Oklahoma, 2011 Oklahoma Alternate Assessment Program (OAAP) mathematics, grades 6–8 ▪ Virginia, 2011 Virginia Grade Level Alternative (VGLA) US history ▪ Tennessee, 2010 Modified Academic Achievement Standards (MAAS) social studies, grades 7–8 ▪ Tennessee, 2010 Achievement Assessments (ACH) reading, grades 5–6 ▪ New Jersey, 2009 Alternate Proficiency Assessment (APA) Language Arts, grades 6–8 ▪ Texas, 2009 Texas Assessment of Knowledge and Skills (TAKS) Alternate science, grades 5, 8, 10, 11 (including general session for all grades/subjects) ▪ Texas, 2009 TAKS Modified General Session for all grades/subjects ▪ Texas, 2008 TAKS Standards Review reading, grades 6–8

Proposed Pearson Psychometric Facilitators		
Name	**Title**	**Standard Setting Experience**
		▪ Texas, 2008 TAKS Modified reading, grades 3–5 (including general session for all grades/subjects) ▪ Texas, 2007 TAKS Alternate mathematics, grades 9–11 (including general session for all grades/subjects) ▪ Texas, 2005 algebra I End-of Course ▪ Michigan, 2005 Michigan Educational Assessment Program (MEAP) social studies, grades 6, 9
Ye Tong	Director, Psychometric Services	▪ Georgia, 2005, EOCT, physical science, high school, lead planner ▪ Michigan, 2006, MEAP, mathematics grades 3–5 ▪ North Carolina, 2006, mathematics grades 3–4 ▪ Georgia, 2007, EOCT US history, high school ▪ New York, 2007, Regents earth science, high school ▪ Georgia, 2008, GHSGT ELA, high school ▪ New York, 2008, Regents Integrated algebra, high school ▪ Georgia, 2009, EOCT mathematics I, high school ▪ New York, 2009, Regents geometry, high school ▪ Georgia, 2010, GHSGT social studies, high school ▪ New York, 2010, Regents algebra II, high school ▪ New York, 2011, Regents English, high school ▪ Georgia, 2011, GHSGT mathematics, high school ▪ Minnesota, 2011, MCA III mathematics 3–8, Modified assessment 5–8 and 10, MCA III Alternate assessment 3–8, lead planner ▪ Virginia, 2011, End of Course Virginia History ▪ Virginia 2012, writing grade 8 ▪ New York, 2013, Common Core assessments grades 3–8 ELA and mathematics, lead planner
James Ingrisone	Research Scientist, Psychometric Services	▪ Virginia, EOC, world history II, high school ▪ Virginia, US history I, middle school ▪ Virginia Grade Level Alternative Assessment (VGLA), Civics and Economics, Middle School ▪ Virginia, EOC, algebra I, high school ▪ Virginia, grade 5 mathematics I, Elementary School ▪ Virginia Modified Achievement Standards Test (VMAST), grade 8 mathematics, junior high school ▪ Virginia Modified Standards Diploma, mathematics, junior high school ▪ Virginia, EOC, reading, high school ▪ North Carolina, EOC, English II, high school
Jason Meyers	Senior Research Scientist	▪ Tennessee, mathematics, grades 5–6 ▪ Kentucky, reading, grades 5–6 ▪ Oklahoma, social studies, grade 8 ▪ Oklahoma, End-of-Course (EOC) US history ▪ Georgia, EOC, Analytic geometry

Proposed Pearson Psychometric Facilitators		
Name	**Title**	**Standard Setting Experience**
Alvaro Arce	Manager, Psychometric Services	▪ Alabama, mathematics, grade 4 ▪ Florida, mathematics, grade 3 ▪ North Carolina, EOC biology, grades 9–11 ▪ Mississippi, biology and history, grade 10 ▪ Tennessee, mathematics, grade 4 ▪ DoDEA, social studies, grades 3 and 6 ▪ DoDEA, science, grades 5 and 7 ▪ Oregon, Spanish reading, grades 3 and 4 ▪ Stanford, Spanish Language Proficiency Test, grade band 3–5 ▪ New York State, English as a Second Language, grades 7–8 ▪ New York City, mathematics Summer test
Marc Johnson	Research Scientist	▪ Arizona, mathematics, grades 3–5 ▪ Arizona, high school, writing ▪ Arizona, English Language Learners, grades 3–5 ▪ Virginia, mathematics alternate assessment, grades 3–5 ▪ Virginia, Modified Achievement Standards assessment, reading, grades 6–8, ▪ Virginia, Modified SOL assessment, reading, grades 5–6 ▪ Kentucky (Lead Facilitator): reading and mathematics grades 3–8, writing grades 5, 6, 8, 10, and 11 ▪ Oklahoma, mathematics, grades 5–6 ▪ Oklahoma, Modified Assessment mathematics, grades 5–6 ▪ Oklahoma, Alternate Assessments Geography grade 7, social studies grade 5, US history grade 8 and high school ▪ Florida, reading, grade 9 ▪ Mississippi, EOC, English II ▪ New Mexico, social studies, grade 11
Tim O'Neil	Manager, Psychometric Services	▪ Wyoming, PAWS writing, grades 7–8 ▪ Oklahoma, Alternate Assessment reading/writing, grades 3–5, ▪ Virginia, EOC, Civics and Economics ▪ Texas, STAAR Modified mathematics, grades 3–8 ▪ ReadyPoint Nursing, College Pediatrics (virtual) ▪ Virginia, reading, grade 7

Proposed Pearson Psychometric Facilitators		
Name	**Title**	**Standard Setting Experience**
Mike Clark	Research Scientist	▪ Oklahoma, mathematics, grades 7–8 ▪ Oklahoma, EOI Modified assessment, US history ▪ Florida, reading, grade 7 ▪ Kentucky, grades 7–8, reading (PLD creation) ▪ Kentucky, writing, grade 10 ▪ Florida, EOC, biology ▪ Arizona, Multiple grades, AZELLA Stage IV (ELL English proficiency exam) ▪ North Carolina, Lead facilitator

Experienced Psychometric Staffing. With Pearson's psychometric staff experience, PARCC will be find a facilitator experienced in a specific type of standard setting event.

In addition to the expert psychometric facilitation that Pearson will provide to lead the standard setting process, Pearson, ETS, and WestEd will provide content facilitators for each panel to provide content expertise and support for discussions.

For each panel, where possible, this content facilitator will be the item development or test development lead of the PARCC project so that the panels have a source to rely on throughout the discussions that have intimate knowledge of the Common Core State Standards, the test design, and the item and test development process. The offered content facilitators from each of the proposed item development vendors are shown in the following figure.

Offered Pearson, ETS, and WestEd Content Facilitators			
Name/Company	**Content Area**	**Panel**	**Standard Setting Experience**
Eric Weisman, Pearson	ELA/Literacy	3 and 4	▪ Texas, ELA, grades 3–8, ▪ Texas, writing, grades 4 and 7 ▪ Texas, reading, English I, II, III ▪ Texas, writing, English I, II, III
Martha Scarborough, Pearson	ELA/Literacy	5 and 6	▪ Texas, reading (SDAA), grades 7–8 ▪ Texas, ELA, grades 9–11 ▪ Texas, reading, English II
Joel Carino. WestEd	ELA/Literacy	7 and 8	None
Kelly King, ETS	ELA/Literacy	9	▪ Stanford Achievement Test, reading/writing, all grades and early reading
Will Steele, ETS	ELA/Literacy	10	▪ Florida Comprehensive Assessment, reading, grades3–10

Offered Pearson, ETS, and WestEd Content Facilitators			
Name/Company	**Content Area**	**Panel**	**Standard Setting Experience**
Chaya Rao, ETS	ELA/Literacy	11	None
Michael Bay-Borelli, Pearson	Mathematics	3 and 4	▪ Texas, mathematics, grades 3–8 ▪ Texas, algebra I ▪ Texas, geometry ▪ Texas, algebra II
Mary Veazey, Pearson	Mathematics	5	▪ American Diploma Project, algebra II, EOC (Arkansas) ▪ American Diploma Project, algebra II, EOC (Ohio) ▪ American Diploma Project, algebra II, EOC (Maryland)
Kellie Taylor-White, ETS	Mathematics	6	None
Christine Reyes-Swank, ETS	Mathematics	7 and 8	▪ Florida Comprehensive Assessment, mathematics, grade 6 ▪ Washington Comprehensive Assessment Program mathematics, grades 3–8 ▪ Texas Educator Certification Test, mathematics 3–8
Ernest Battle, ETS	Mathematics	Algebra I/ Integrated I	▪ Tennessee Gateway and EOC, algebra and mathematics grades 3–8
Will Wanamaker, ETS	Mathematics	Geometry/ Integrated II	None
Luis Saldivia, ETS	Mathematics	Algebra II/ Integrated III	▪ PPAA in Puerto Rico grades 3–5 ▪ California across different grades ▪ CLEP college algebra

Breadth of Pearson Team. The combined Pearson team can provide an experienced content facilitator with detailed knowledge of the applicable Common Core State Standard.

References

Haertel, E. H., Beimers, J. N., & Miles, J. A. (2012). The briefing book method. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 283–299). New York, NY: Routledge.

McClarty, K. L., Way, W. D., Porter, A. C., Beimers, J. N., & Miles, J. A. (2013). Evidence-based standard setting: Establishing a validity framework for cut scores. *Educational Researcher, 42*(2), 78-88.

O'Malley, K., Keng, L., & Miles, J. (2012). From Z to A: Using validity evidence to set performance standards. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 301–322). New York, NY: Routledge.

Requirement

V.E.3.C. Vertically Moderated PLDs

Response

Refer to the response to V.E.2.Performance Level Descriptors for Standard Setting that fully covers all PLD requirements including approach to keeping the revised PLDs vertically moderated across grades within each content.

c) Description of approach in ensuring that the revised PLDs are vertically moderated across all grades within each content area.

V.E.4. Participant Recruitment & Selection

Requirement

Response Requirements for Section V.E.4.

a) Description of the approach and procedures to complete all the responsibilities/tasks specified in Section V.E.4

Deliverables for Section V.E.4.

a) Plan outlining the contribution of each specific state in each panel by panelist background, updated as needed based on actual recruitment
b) Document listing the final panelists for each panel by state and background.

Response

Participant Recruitment and Selection

Pearson will work with the PARCC member states to provide that each of the 12 standard setting panels (shown in the following figure) are comprised of 20 panelists that reflect the diversity of the state with regards to gender, ethnicity, demographic locations (rural, urban, suburban), and other demographic variables as named by the Partnership.

Panel Distribution Across Subject Grade, and Course		
Panel	**Subject/Grade/Course**	
	Mathematics	**ELA/Literacy**
Panel 1	grade 3 and 4	
Panel 2	grade 5 and 6	
Panel 3	grade 7 and 8	
Panel 4	algebra I and integrated I	
Panel 5	geometry and integrated II	
Panel 6	algebra II and integrated III	
Panel 7		grade 3 and 4

Panel Distribution Across Subject Grade, and Course		
Panel	Subject/Grade/Course	
	Mathematics	ELA/Literacy
Panel 8		grade 5 and 6
Panel 9		grade 7 and 8
Panel 10		grade 9
Panel 11		grade 10
Panel 12		grade 11

Recruiting Diverse Panels. The Partnership can rely on Pearson to recruit panels that reflect the diversity and variables named by the Partnership.

The goal will be for each panel to include teachers of special populations such as special education, English language learners and talented/gifted, and others deemed relevant to reflect the performance expectations of the member states.

Based on the provided requirements, different configurations required for three subsets of the panels:

1. Grade 3–8 panels must have the following composition:
   - 12 grade-level teachers
   - Two above-grade teachers
   - Six other stakeholders (e.g., parent, business community)
2. Grade 9–10/algebra I/geometry/integrated math I and II panels must have the following composition:
   - 10 grade-level teachers
   - Two above-grade level teachers
   - Six other stakeholders (e.g., parent, business community)
   - Two non-teacher educators
3. Grade 11/algebra II/integrated math III panels must have the following composition:
   - Eight grade-level teachers
   - Six higher education faculty and staff
   - Two technical/vocational educators
   - Four other stakeholders (e.g., parent, business community)

To meet the requirements for the composition of the panels and reflect the diversity of the PARCC member states, it will be necessary for Pearson to work closely with the states to understand the composition of their student population to provide representativeness of the panelists at the standard setting event.

In conjunction with the PARCC member states, Pearson will develop high-level targets to hit in terms of the major subgroups (gender, ethnicity, and special populations, for each of the three panel configurations based on data available at http://nces.ed.gov/nationsreportcard/states (which is based on the Common Core of Data survey). This data will provide a snapshot of registration information for each state that can be used as targets for recruiting panelists.

Additionally, targets based on geographic region will need to be provided by the PARCC member states unless the available urban-rural data from the latest U.S. Census data (retrieved on 12/10/2014 from www.census.gov/compendia/statab/2012/tables/12s0029.xls) is considered appropriate.

The following figure shows a mock target table that would be created for each state showing their specific demographic targets. This example shows data for a real state with a student population with a race/ethnicity breakdown of 82 percent white, 5 percent black, 9 percent Hispanic, and 2 percent Asian; a student program breakdown of 14 percent of students with Individualized Education Plans (IEPs) and 4 percent in Limited English Proficiency programs; and a 64 percent urban and 36 percent rural breakdown.

Mock Table for Demographic Target

Panelist Representative	Race/Ethnicity				Gender		Population				
	African-American	Asian	Caucasian	Hispanic/Latino	F	M	Rural	Urban	REG	SWD	ELL
Grade-Level (12)	1	1	9	1	6	6	4	8	10	1	1
Above Grade-Level (2)			2		1	1	1	1	2	0	0
Other–Parent (3)	1		2		1	2	1	2	2	1	0
Other–Community (3)			2	1	2	1	1	2	2	1	0
Percent	10%	5%	75%	10%	50%	50%	35%	65%	80%	15%	5%
Total	**20**				**20**		**20**		**20**		

Mock Target Table. The Pearson team will work with the Partnership to populate standard setting panels to reflect your demographic preferences and balance policy considerations.

As can be seen from the figure, it may be necessary to oversample certain groups (e.g., parent or community members of outside of the Caucasian group) to obtain a variety of expectations upon which policy considerations can be made. This determination will be made in conjunction with the PARCC member states.

Once the demographic targets are agreed upon by the PARCC member states, the PARCC member states will solicit volunteers and provide a list of volunteers and their demographic information to Pearson so that Pearson and the PARCC member states can work together to select panelists to meet targets as efficiently as possible.

It should be noted that challenges are likely in reflecting the diversity of the demographics of the student population in the selection of the educator panelists because often teacher diversity does not reflect student diversity (Boser, 2011). Additionally, trying to hit all targets with such a small panel may require that the PARCC states provide guidance on which targets are most critical to hit.

References

Boser, Ulrich (2011). Teacher Diversity Matters: A State-by-state analysis of teachers of color. Center for American Progress. Retrieved on 12/10/14 from http://www.americanprogress.org/wp-content/uploads/issues/2011/11/pdf/teacher_diversity.pdf.

V.E.5. Special Studies to Inform Standard Setting

Requirement

Response Requirements for Section V.E.5.

a) Description of the approach and procedures proposed to complete all the responsibilities/tasks specified in Section V.E.5

b) Description of the nature of coordination required with other PARCC Contractors and with the third party organizations in conducting each study.

c) Description of general approach and the rationale for the

   i. study design and data analyses including (but not limited to) approach and rationale for selection of items and tasks to be included in the studies, and data collection tools and methods if and where applicable

   ii. sampling frameworks (including sample size) and procedures for recruiting and securing required sample sizes if sampling is required

d) The general approach described by the Offerors in the proposal will be the basis for a detailed study plan to be required at a later time from the contracted vendor.

e) Plans for mitigating the negative impact of missing data, where appropriate.

f) Description of quality control procedures to ensure accuracy in data processing.

The general approach described by the Offerors in the proposal will be the basis for a detailed study plan to be required at a later time from the Contractor.

Deliverables for Section V.E.5.

a) Memoranda of Understanding with third parties whose collaboration is needed in standard setting research studies

b) Draft and final Study Plan for each study seven weeks in advance of the start of the study (as described in Table V.E.1). The plan shall include timelines, study design, sampling specifications, and data analysis methods

   i. The Contractor shall present the study plans to the PARCC Technical Advisory Committee (TAC), one or more PARCC Operational Working Groups (OWG), and/or other expert reviewers for feedback. The Contractor shall incorporate such feedback.

c) The final study plan three weeks in advance of the start of the study (as described in Table V.E.1).The final plan shall be approved by PARCC before the study is conducted.

d) Draft and final data collection instruments, survey and/or data coding schemas three weeks in advance of the start of the study (as described in Table V.E.1).

i. Contractor shall provide all data collection instruments (e.g., interview protocols, observation protocols, surveys, assessment forms) and coding schemas to the Partnership Manager in draft form. Each data collection instrument and coding schema will be reviewed by the TAC, one or more PARCC OWGs, and/or other expert reviewers. The Contractor shall revise the data collection instruments and coding schema accordingly, prior to use in the study.

e) All data (raw, or scored, or coded or processed) collected and processed for each study in a digital format specified by PARCC upon execution of the contract.

f) Draft study reports within six weeks of the completion of data collection or a date mutually-agreed upon by the Partnership and the Contractor.

i. Draft reports for each study will be reviewed by the TAC, one or more PARCC OWGs, and/or other expert reviewers for feedback.

ii. Draft reports shall indicate the Principal Investigator, data analysts and Contractor staff who reviewed and approved the submission of the draft report

g) Final study report within two weeks after the feedback for the draft report is provided or a date mutually-agreed upon by the Partnership and the Contractor

i. The final report must include the theoretical framework and design rationale that cites relevant, peer-reviewed published work and, when appropriate, unpublished technical reports; how the study addressed relevant standards in the Standards for Educational and Psychological Testing (1999); and how the results contribute to the body of evidence to support the valid interpretation of scores.

ii. The final report shall include an executive summary of results, and specific recommendations of action

iii. Final reports shall indicate the Principal Investigator, data analysts and Contractor staff who reviewed and approved the submission of the final report

h) An action plan based on recommendations in the final study report

i. The action plan shall indicate responsibilities of each party involved along with a timeline for each action

ii. The Contractor shall manage the action plan and provide update reports for actions that require Contractor follow up or involvement based on a schedule proposed by the Contractor and mutually-agreed upon by the Partnership and the Contractor

V.E.5.A. Study 1: Benchmark Study to Inform PARCC Middle and High School Performance Standards

Response

Special Studies to Inform Standard Setting

Because the interpretation of performance at PARCC Level 4 has specific linkages to college success (i.e., approximately 0.75 probability of earning a C or better in relevant credit-bearing college courses), there is an ongoing need to align the performance standards to external indicators. The extent to which the standards mirror external indicators lays the foundation for validating their interpretations.

Benchmark Studies

Pearson will conduct a series of studies to inform reasonable ranges where the Level 4 cut score may be placed for grade 4 ELA/Literacy and mathematics, grade 8 ELA/Literacy and mathematics, grade 11 ELA/Literacy, Algebra II, and Integrated Mathematics III. The following table provides a list of benchmarks that will be considered for inclusion in the final study report for each PARCC assessment.

More detailed descriptions of the benchmarks, the rationale for including them, and the specific data source(s) are provided in the sections that follow.

PARCC Assessments: Benchmarks to be Considered for Final Study Report						
Comparison Assessments & Benchmarks	Grade 4 ELA/Literacy	Grade 4 Mathematics	Grade 8 ELA/Literacy	Grade 8 Mathematics	Grade 11 ELA/Literacy	Algebra II & Integrated Mathematics III
NAEP	4 Reading Proficient	4 Mathematics Proficient	8 Reading Proficient	8 Mathematics Proficient	12 Reading Proficient	12 Mathematics Proficient & at or above 163
ACT			EXPLORE Reading (16) & English (13)	EXPLORE Mathematics (17)	ACT Reading (22) & English (18) Compass Reading (89) & Writing (77)	ACT Mathematics (22) Compass Algebra (52)
College Board			ReadiStep Critical Reading & Writing	ReadiStep Math	SAT Critical Reading (500) & Writing (500)	SAT Mathematics (500)
International	PIRLS			TIMSS	PISA Reading Literacy	PISA Mathematics Literacy
SBAC	4 Reading Level 3	4 Mathematics Level 3	8 Reading Level 3	8 Mathematics Level 3	11 ELA Level 3	11 Mathematics Level 3
ASVAB					AFQT ACT English + Mathematics (48 for Army, 45 for Marines) ACT English (22 for Navy) SAT combined (1100 for Army, 1000 for Marines) SAT Reading (530 for Navy)	AFQT ACT English + Mathematics (48 for Army, 45 for Marines) ACT English (22 for Navy) SAT combined (1100 for Army, 1000 for Marines) SAT Reading (520 for Navy)
State Assessments	NY, TX, MI, VA 4 Reading	NY, TX, MI, VA 4 Math	NY, TX, MI, VA 8 Reading	NY, TX, MI, VA 8 Math	NY Regents TX STAAR MI Merit Exam VA EOC	NY Regents TX STAAR MI Merit Exam VA EOC

NAEP Assessments

Results from the 2013 National Assessment of Educational Progress are currently available for grades 4 and 8 reading and mathematics, and the results for grade 12 will be available in 2014. A recent NAEP report (Fields, 2013) indicates that the percentage of students scoring at the Proficient level or higher on the grade 12 reading assessment, or earning a score of 163 or higher on the grade 12 mathematics assessment (which is below the current grade 12 mathematics Proficient cut score of 176), is a plausible estimate of the percentage of students possessing the knowledge, skills, and abilities to make them academically prepared for college. Because this definition is similar to PARCC's college- and career-ready definition, it provides a reasonable point of comparison. For mathematics, comparisons can also be made to the percentage of students scoring Proficient or higher on NAEP. No specific linking studies have been done for NAEP in grades 4 and 8, but given some similarities in the level of knowledge and skills required by the PLDs of each grade level (e.g., grade 4 proficient vs. grade 8 proficient), it seems reasonable to use the percentage of students scoring Proficient or higher on the grades 4 and 8 assessments as one piece of information to inform the reasonableness of the Level 4 cut score on the PARCC assessment[1]. For grades 4 and 8, results are published for each state, so analyses can be conducted for the nation as well as for the subset of PARCC states.

ACT and College Board Assessments

Both ACT and the College Board (for the SAT test) annually publish the percentage of students meeting their college readiness benchmarks. This information is generally available at the national level and for individual states. However, for most states, the results reflect the sample of students who elected to take the assessments and are not representative of the whole state. This can lead to biased estimates of the percentage of students who are college ready. For this study, we will consider results nationally and for those PARCC states where the majority of students take the ACT or SAT assessments. Although the ACT definition of college readiness (i.e., 50 percent likelihood of earning a B or better and 75 percent likelihood of earning a C or better) is a bit closer to the PARCC definition of college- and career-readiness than the College Board's definition (i.e., 65 percent probability of earning a freshman year GPA of B- or higher), both may be informative. In addition, the College Board has previously released information about the relationship between their individual content-area assessments (i.e., SAT critical reading and mathematics) and success in particular college courses for use in standard setting by other states (e.g., Texas, New York). PARCC may be able to request to use the same information in order to have results that more closely match their definition. This may also require states to request state-level data from the College Board, as publically released results are typically average scores and percentage of students meeting the benchmarks rather than score distributions. In addition, the current published benchmarks for ReadiStep (11.8) reflect a composite of scores. PARCC will need

[1] Note that the percentage of student scoring Proficient or higher may be an upper bound given that the Grade 12 Math preparedness cut score is lower than the grade 12 Proficient cut score.

to work with the College Board to determine appropriate benchmarks for comparison on the individual subject tests.

International Assessments

Results from international assessments may also be relevant for PARCC standard setting. In 2012, the National Center for Education Statistics (NCES) conducted a study to link NAEP with the Trends in International Mathematics and Science Study (TIMSS). In October 2013, results were released which predicted the percentage of students in each TIMSS performance level for all U.S. states based on this linking (NCES, 2013). The percentage of students across the nation and in PARCC states reaching the TIMSS performance level that is most aligned with PARCC Level 4 will be estimated. Results from the Program for International Student Assessment (PISA) administration of 2009 could also be considered including the percentage of students in the U.S. who reached each of the seven performance levels and the performance level most aligned with PARCC Level 4 for the reading and mathematics literacy components (Fleischman, Hopstock, Pelczar, Shelley, & Xie, 2010). The results for PISA 2012 will be released December 2013 and will also be included in this benchmark study. In addition, PARCC is conducting its own study with linkages between PARCC, TIMSS 2011, and PISA 2012 with results planned to be available in September 2014. Although that PARCC linking study will not be part of the benchmarking report, the results from that study will be incorporated in the standard-setting process along with the other study information that is gathered. Finally, data from the 2011 Progress in International Ready Literacy Survey (PIRLS) can be used to inform the performance standards for PARCC grade 4 reading. National results estimating the percentage of students in each of the five performance levels are available on the NCES website. The percentage of students at or above the performance level that best aligns with PARCC Level 4 can be estimated.

SBAC Assessments

Because the Smarter Balanced Assessment Consortia is targeting standard setting for summer 2014, it is unlikely that information about the percentage of students in each performance level will be available for inclusion in this benchmarking report (due August 2014). The information should be available, however, prior to the PARCC operational standard-setting meeting in summer 2015 (and possibly for the field-trial in spring 2015) and can be incorporated as part of the information provided to panelists in the standard-setting process. Comparison benchmarks can include the percentage of students predicted to be in SBAC Level 3 or higher (prior to 2015 operational assessment) or the percentage of students in Level 3 or higher (after 2015 results are released). The Level 3 cut score for SBAC is likely best aligned to the Level 4 cut score for PARCC.

Armed Services Vocational Aptitude Battery

The Armed Services Vocational Aptitude Battery (ASVAB) is a multi-aptitude battery consisting of 10 subtests. For applicants seeking general military enlistment, The Armed Forces Qualifying Test (AFQT) scores are used to determine eligibility. The AFQT is a subset of the ASVAB, comprised of scores on Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Mathematics Knowledge.

AFQT scores are expressed in percentile units and were normed based on data from the Profile of American Youth (PAY97) project. Cuts have been established separately for each branch of the armed services and range from 31 for Army to 40 for Coast Guard. This means that nationally, 60-69% of 18-23 year-olds possess the requisite knowledge and skills in ELA and mathematics to enlist in the military. The PARCC assessments, however, measure students' knowledge and skills in grade 11. According to research using the ASVAB for the career explorations program, grade 11 students do not perform as well on average as the norming sample (Defense Manpower Data Center, 2004). In addition, the military has cut scores for eligibility to join the Reserve Officers' Training Corps (ROTC). These minimum eligibility thresholds are more stringent than the AFQT cuts and are based on college admissions tests—the SAT and ACT. The percentage of students meeting each of these cut scores could also be considered in the PARCC standard-setting process. Although benchmark studies of military performance can be conducted, the linkage between the interpretation of PARCC Level 4 performance and eligibility for enlisting in the military or enrolling in ROTC may be weak.

NAEP Preparedness Research Studies

The recommendations listed under the first benchmark (NAEP assessments) include considerations based on recently published reports from NAEP (Fields, 2013) and their reporting plan which was approved August 2013. Results will be published in 2014, included in the benchmark report for this study. In addition, the NAEP preparedness research will be reviewed and can inform the process of conducting other studies in support of PARCC standard setting.

State Assessments

Several states have recently set performance standards indicative of college- and career-readiness and have incorporated empirical data as part of the standard-setting process. Research from standard-setting meetings in New York, Texas, Michigan, and Virginia (as well as other states identified as relevant) will be reviewed for (1) empirical research results that could be informative or relevant to PARCC and (2) percentage of students achieving (or estimated to achieve) the performance level that most closely aligns with PARCC Level 4.

Literature Review of Existing Reports

Pearson's researchers will conduct a review of the literature to gather relevant statistics on graduation, remediation, workforce preparedness rates, analyses of post-secondary course content and learning materials, post-secondary course performance, and other relevant statistics including results from existing studies and surveys from PARCC states or at the national level.

Statistical Methods

For each of these studies, analyses will involve the percentage of students reaching a particular benchmark score on each external assessment. For example, we will report the percentage of students overall and for the PARCC states who reached Proficient or higher on NAEP grade 4 reading. These percentages will inform the percentage of students who would reasonably be classified in Level 4 or higher on the various PARCC assessments. There will be no direct linkages made between the external assessments and the PARCC theta or scale score scale as part of this benchmarking study. The results of each study will be contextualized according to several considerations of the quality and relevance of each comparison point. For example, the representativeness and motivation level of the examinees are important considerations when evaluating the utility of each comparison benchmark.

Data to Inform Vertical Moderation

PARCC may also consider using external assessment information to inform the vertical moderation process—that is the backward linking of performance standards from high school down through elementary school. Using existing state or commercial data built with vertical scales, the shape of increases in student performance can be evaluated. For example, students may make larger gains in reading comprehension at lower grade levels (as evidenced by a steeper trajectory in appropriate text complexity levels at lower grades, see Williamson, Fitzgerald, & Stenner, 2013), so the distance between performance standards may also be greater at lower grade levels.

References

ACT (2013, September) *What are the ACT college readiness benchmarks?* Iowa City, IA: ACT. Retrieved from: http://www.act.org/research/policymakers/pdf/benchmarks.pdf

Defense Manpower Data Center (2004). *ASVAB norms for the career exploration program.* Retrieved from http://www.asvabprogram.com/downloads/asvab_norms.pdf

Fields, R. (2013, November). *Validity Argument for NAEP Reporting on 12th Grade Academic Preparedness for College.* Washington, DC: National Assessment Governing Board.

Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., Shelley, B. E., & Xie, H. (2010). *Highlights from PISA 2009: Performance of U.S. 15 year-old students in reading, mathematics, and science literacy in an international context.* Retrieved from: http://nces.ed.gov/pubs2011/2011004.pdf

National Center for Education Statistics (2013). *U.S. states in a global context: Results from the 2011 NAEP-TIMSS linking study.* Retrieved from: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2013460

Williamson, G. L., Fitzgerald, J., & Stenner, A. J. (2013). The Common Core State Standards' text complexity trajectory: Figuring out how much complexity is enough. *Educational Researcher, 42*(2), 59-69.

Requirement

V.E.5.C. Study 3: Postsecondary Educators Judgment Study to Inform Cut Scores in PARCC High Schools Assessments

Response

Post-Secondary Educator Judgment Study

Pearson proposes to gather recommendations from postsecondary educators including vocational institutions regarding the minimum level of performance on PARCC items that would indicate students were academically ready to take and succeed in the postsecondary courses under consideration. Postsecondary courses to be considered for Study 3, as of this writing, include: College English Composition, Literature, and technical courses requiring college-level reading and writing; and College Algebra, Introductory College Statistics, and technical courses requiring an equivalent level of mathematics (based on PARCC's College- and Career-Ready Determination Policy[2]).

Given the focus on college- and career-readiness, judgment studies utilizing vocational stakeholders and college faculty as subject matter experts can provide important information that is useful to the context of setting standards related to college- and career-ready goals (level 4 as defined by the PARCC CCR policy). Pearson has experience in supporting large-scale assessment programs with collecting postsecondary stakeholder judgments concerning college- and career-readiness in support of standard-setting activities in Virginia, Texas, and the American Diploma Project. The following figure outlines the purpose, sample, data sources, and possible analyses for Study 3.

Study 3: Postsecondary Educators Judgment Study to Inform Cut Scores in PARCC High School Assessments	
Purpose	▪ Gather recommendations from instructors and/or professors teaching relevant courses (College English Composition, Literature, College Algebra, Introductory College Statistics and technical courses requiring reading, writing or mathematics) on cut scores for PARCC ELA/Literacy Grade 11 and Algebra II that separate students who are academically ready for the relevant courses under consideration from those who are not ○ Academically ready will defined as students who will earn a course grade of *C* or better. ▪ Data from the study will be used in standard setting to inform the college- and career-readiness cut (Level 4)

[2] http://www.parcconline.org/sites/parcc/files/PARCCCCRDPolicyandPLDs_FINAL_0.pdf.

Study 3: Postsecondary Educators Judgment Study to Inform Cut Scores in PARCC High School Assessments	
Sample	▪ 300 instructors and/or professors ○ Currently teaching College English Composition, Literature, College Algebra, Introductory College Statistics and technical courses requiring reading, writing or mathematics Enrolled in College English Composition, Literature, College Algebra, Introductory College Statistics or technical courses requiring reading, writing or mathematics ○ Selected from a Partnership provided list of campuses, representing two-year and four-year institutions (both open and selective registration) as well as institutions of vocational or technical instruction selected
Data Sources	ELA/Literacy Grade 11 ▪ Item-level evaluations of whether or not an academically prepared student would get ELA/Literacy Grade 11 EOY and PBA PARCC items correct on day 1 of instruction in the relevant course ▪ Instructor/professor demographics (state, school, course being taught, number of years teaching relevant course, highest degree held, gender, ethnicity) Algebra II ▪ Item-level evaluations of whether or not an academically prepared student would get Algebra II EOY and PBA PARCC items correct on day 1 of instruction in the relevant course ▪ Instructor/professor demographics (state, school, course being taught, number of years teaching relevant course, highest degree held, gender, ethnicity)
Potential Analysis Method(s)	▪ Descriptive statistics of the cut scores resulting from item-level judgments disaggregated by subgroups ○ Relevant Course (College English Composition, Literature, College Algebra, Introductory College Statistics, technical equivalents) ○ Institution Type (technical, two-year, four-year (open), four-year (selective)

The following figure shows the type of institution (Two Year/Technical, 2YT; Four Year Open, 4YO; and Four Year Selective, 4YS), the course type to be included (Composition, Comp; Literature, Lit; Technical Equivalent, TE; College Algebra, Algebra; and Introductory Statistics, Stats) and the number of respondents to be recruited for each PARCC assessment to be evaluated (ELA/Literacy Grade 11 or Algebra II) for each institution type.

Sampling Plan: Number of Participants per Institution Type and PARCC Assessment							
	PARCC Assessment						
	ELA/Literacy Grade 11			Algebra II			
Institution Type	**Comp**	**Lit**	**TE**	**CAlg**	**Stats**	**TE**	**Total Participants**
2YT	10	10	30	10	10	30	100
4YO	25	25	n/a	25	25	n/a	100
FYS	25	25	n/a	25	25	n/a	100
Total Participants	**60**	**60**	**30**	**60**	**60**	**30**	**300**

Study 3 Process Overview

Study 3 participants will be provided with access to a Pearson system (ABBI: Asset Banking & Building for Interoperability; see V.A. Test Development, for a description of the system) which will allow secure access to the items of interest for this study. The items selected for use in this study will be those administered on intact EOY and PBA forms as part of the PARCC Field Test administration. Selecting intact forms will allow the results of the study to be more easily placed on the theta metric for standard-setting purposes as well as provide a selection of items that meets the operational blueprint.

Inside the ABBI system, participants will be able to access a static view of the selected items on a review screen as well as see associated passages, relevant item-level meta-data, and scoring rubrics for constructed-response items. Items will be provided in the field-test sequence. If panelists wish to see an interactive version of the technology-enhanced items, they will be able to click a button which will open the item in the PARCC Assessment System platform. Voting buttons and comment fields within the ABBI system can be configured to capture participant judgments for data analyses purposes.

Participants will be provided with approximately 1.5 hours of web-based training on the purpose of the study, background of the PARCC assessments, the meaning of performance level 4, and training on using the ABBI system. In order to support the varying schedules of the participants, this training will be provided on three different occasions with participants selecting the training event that works best for them. The participants will be able to access the system and input their judgments during set windows that will allow them to work at their own pace and according to their own daily schedule. As part of the training, participants will be instructed to review each item in the selected forms and determine the number of points an “academically ready” student would earn on that item if the student encountered it on the first day of instruction in the participant’s college-level course. For the purposes of this study, “academically ready’” will be defined as able to earn at least a *C* or its equivalent in the college-level credit-bearing course without remediation and/or other criteria that the Partnership may specify at the time of the study.

Panelists will begin with the first item in the field-test forms selected for the study and answer the question, “Would an ‘academically ready’ student get this item correct on their first day in my class?” If the answer is “yes,” they will be instructed to select a “1” using the voting buttons for the item; if the answer is “no,” they will be instructed to select a “0” using the voting buttons for the item. The panelists will work through the field-test forms selected for the study in this fashion until they encounter an open-ended item. For open-ended items, the question asked of themselves changes to, “How many points would an ‘academically ready’ student receive if they answered this question on their first day in my class?” Once they review the provided scoring rubric for the item in question and determine a score, they will be instructed to use the voting buttons to select the appropriate score. They will proceed in this manner until all the items on the selected field test forms have been reviewed and judged.

Study 3 Analysis Methods Overview

The data from Study 3 will be analyzed by summing up the number of points participants judged to be necessary for a student to be considered "academically ready" on the form of interest and aggregated based on the type of institution (two-year/technical, four-year open registration, and four-year selective registration) and the courses being taught by the participants (composition, literature, college algebra, introductory statistics, and technical equivalents). The data from these aggregated cut scores can then be included in materials prepared for Study 4, as appropriate, for the evaluation of the proposed standard-setting process. An example layout of the descriptive statistics that could be provided as a result of Study 3 is shown in the following figure.

Example Results Table Summarizing Cut Score Recommendations from Study 3							
	College-Level Course	**Institution Type**	**Mean**	**Median**	**Min**	**Max**	**Std Dev**
ELA/Literacy Grade 11	College English Composition	Four-Year Open					
		Four-Year Selective					
		Combined					
	Literature	Four-Year Open					
		Four-Year Selective					
		Combined					
	Technical Equivalent	Two-Year/Technical					
Algebra II	College Algebra	Four-Year Open					
		Four-Year Selective					
		Combined					
	Introductory Statistics	Four-Year Open					
		Four-Year Selective					
		Combined					
	Technical Equivalent	Two-Year/Technical					

Additional analyses can be done using participant demographics (such as gender, ethnicity, number of years teaching the college-level course under consideration, type of degree held) if the Partnership desires. One way to display the range of cut scores resulting from this judgment study is illustrated in the following figure where a raw score of 81 might represent the aggregated cut score of the Composition instructors/professors participating from two-year or technical ("2YT") institutions for students to be considered 'academically ready' while a raw score of 117 would represent the cut score for Literature instructors/professors participating from four-year (open, "4YO") and four-year (selective, "4YS").

Example Range of Cut Scores from Judgment Study Analyses		
PARCC Raw Score	PARCC Theta	Judgment Studies Cut Scores
0	-4.091	
⋮	⋮	
25	-3.551	
26	-3.181	
27	-3.056	
28	-2.682	
29	-2.456	
⋮	⋮	
70	0.478	
71	0.514	
72	0.523	
73	0.542	
74	0.633	
75	0.794	
76	0.898	
77	0.898	
78	1.024	
79	1.128	
80	1.204	
81	1.454	2YT Composition
82	1.756	
83	1.808	2YT Technical
84	2.068	2YT Literature
85	2.309	
⋮	⋮	4YO Composition
110	2.432	
111	2.812	4YS Composition
112	2.820	
113	2.836	
114	2.952	
115	2.973	
116	3.023	
117	3.159	4YO & 4YS Literature
118	3.404	
⋮	⋮	
130	3.719	
131	3.728	
132	3.733	
133	4.633	
⋮	4.678	
150	4.997	

One additional area of information that can be pursued with the data collected in Study 3 is to rank order the evidence statements assessed by the number of points assigned by the participants. This type of information could be useful during standard-setting activities to provide a higher education perspective to the content-based discussions at the Level 4 cut score as to what evidence statements are ranked higher (in terms of student mastery being necessary for success) by a majority of higher education instructors/professors in order to be ”academically ready.”

This information could be very useful at the standard setting by providing empirical data about which evidence statements are ranked most important for supporting “readiness.” These data could help anchor the standard setting panelists’ discussions related to which evidence statements should be included in the threshold descriptions for the Level 4 cut score. Threshold descriptions reflect the expectations of what a “bubble” or “just-barely Level 4” student should be able to demonstrate.

Requirement

V.E.5.E. Study 5: Longitudinal Study of External Validity of PARCC Performance Standards

Response

Longitudinal Study of External Validity of the PARCC Performance Standards

The purpose of this study is to directly support the validity of the PARCC college- and career-ready performance levels for Grade 11 ELA/Literacy, Algebra II, and Integrated Mathematics III. Specifically, students who complete the grade 11 PARCC assessments in spring 2015 will be matched with their ACT and/or SAT scores (likely taken in a similar timeframe) and with their course performance in entry-level classes at post-secondary institutions.

Data Sources

Data for this study will come from state longitudinal data systems (SLDS), ACT, and the College Board. Pearson can work with states to request data from the 2015 and 2016 administrations of the ACT and SAT assessments in cases where the state does not already receive this data. Information on post-secondary course performance will be provided through existing SLDS for each state. According to a recent study by the Data Quality Campaign (http://www.dataqualitycampaign.org/your-states-progress/), the majority of states currently include their K-12 and post-secondary data. Of the 19 states currently listed on the PARCC website, 17 connect this data annually (only Arizona and Ohio do not). Therefore, it seems that most states would have this data, and Pearson will work with them to extract relevant pieces for the purposes of this study. Using this existing data source will allow for a large number of students to be included in the study, although linkages in SLDS tend to be limited to students that remain within the state. We could work with PARCC states to develop agreements to share higher education data across state lines in order to match more students' high school PARCC results to their post-secondary performance. Students who take the grade 11 PARCC assessments in spring 2015 should enter college and complete their first semester of courses in fall 2016. In 2017, data could be extracted from the SLDS and used for the purposes of this study.

Analysis Methods

The first data available will be the linkage between PARCC grade 11 assessments and ACT and SAT. The purpose of this analysis is to relate the rigor of performance at PARCC Level 4 to the ACT and SAT college readiness benchmarks. One measure of rigor is simply the percentage of students meeting each achievement level. If students are reaching the ACT benchmark, but the same students are not meeting Level 4 on PARCC, the PARCC assessment could be interpreted as more rigorous. In addition, regression analyses can be conducted to predict the most likely ACT and SAT score for a student just meeting the Level 4 standard on PARCC.

Likewise, PARCC performance can be predicted for students just meeting the ACT or SAT college readiness benchmarks. Taken together these analyses can provide a picture of the relative rigor of the PARCC, ACT, and SAT assessments.

The second part of the study involves analysis of post-secondary registration and performance data. PARCC Level 4 should indicate approximately 0.75 probability of earning a C or better in entry-level, credit-bearing courses. The post-secondary data will be coded to determine (1) whether students enrolled in relevant credit-bearing courses, and (2) what grade they earned in each course (whether the grade was C or better). Using logistic regression (where 1 = earned a C or better in a credit-bearing course), we can determine the likelihood that a student reaching Level 4 will earn a C or better—and how the likelihood compares with the 0.75 probability included in the college- and career-ready determination policy.

A second analysis we will conduct to evaluate alternate cut score options, using the post-secondary data, is receiver operating characteristic (ROC) analysis. The prediction accuracy of PARCC's Level 4 cut score and possible alternate cut score options will be assessed using the ROC analysis tools of sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV), which are described in the following figure.

Evaluate Alternate Cut Score Options			
	Not Successful (Remediation, Low Grades)	**College Success (C or better)**	**Predictive Value**
PARCC Level 1, 2, or 3	True Negative (TN)	False Negative (FN)	$NPV = \frac{TN}{TN + FN}$
PARCC Level 4 or 5	False Positive (FP)	True Positive (TP)	$PPV = \frac{TP}{FP + TP}$

Sensitivity/ Specificity	$Specificity = \frac{TN}{TN + FP}$	$Sensitivity = \frac{TP}{FN + TN}$	

As displayed above, the sensitivity at a specific cut score is the proportion of students who were successful in college who scored at least Level 4 on the PARCC assessment. The specificity of a specific cut score is the proportion of students not successful in college who did not reach Level 4 on the PARCC assessment. The PPV associated with a cut score describes the probability that students were successful in college given performance at Level 4 or higher on the PARCC assessment, and the NPV describes the probability that students were not successful in college given performance below Level 4 on PARCC.

These can be useful in evaluating the types of prediction errors made using the original PARCC cut scores and for PARCC to consider whether an alternate cut score might better align with the consortium goals.

For example, the higher the PPV values associated with a cut score, the more certain stakeholders can be that performing at Level 4 on PARCC means the student will be successful in entry-level, credit-bearing college courses in the future. In contrast, the higher the sensitivity values a cut score produces, the more likely it is that the students who will ultimately be successful in entry-level college courses are meeting the Level 4 performance standard. The ROC analysis can help PARCC evaluate trade-offs in prediction at different cut score values as well as determine the cut score that best balanced false positives and false negatives (if this is a goal of the consortium).

Although this type of analysis is typically used in the medical literature, it can be useful in helping PARCC evaluate the predictive validity of the Level 4 cut score and weighing the relative importance of various types of prediction errors. Using this type of analyses, PARCC could make claims such as, “Students who meet Level 4 performance on PARCC go on to earn a grade of C or better in entry-level college courses 75 percent of the time.” This claim is directly associated with the PPV.

Data Security

Regulatory requirements are set forth by federal statutes such as the Federal Educational Rights and Privacy Act (FERPA**), The Health Insurance Portability and Accountability Act** (HIPPA) and the Protection of Pupil Rights Amendment (PPRA) demand secure handling of confidential data.

During our many years of processing private information, Pearson has developed standards for maintaining the security of confidential data. By aligning our security policies and principles with ISO/IEC 7002, our protection strategies adhere to internationally recognized standards and best practices in security.

Pearson utilizes security methods to prevent unauthorized access to data. For example, we will provide:

1. Names of Pearson Personnel that will access the data
2. Description of the storage procedures for the data

In addition, we will keep the data safe by using industry practices for securing the data. A multi-layered security strategy includes encrypted network traffic; certificate-basis authentication (issued by VeriSign); application-level, role-based authorization and entitlements; and physical, network, and procedural security.

Transfer of state data will use our Secure File Transfer Protocols (SFTP) sites or other methods with equivalent encryption and protection. Upon study completion, Pearson will provide all data collected and processed for the study to PARCC in a digital format.

V.E.6. Standard Setting Design and Implementation

Requirement

V.E.6.A. Tools for Presenting Materials, Data Processing and Data Reporting
V.E.6.B. Design

Response Requirements for Section V.E.6.

a) Description of the approach and procedures to complete all the responsibilities/tasks specified in Section V.E.6
b) Description of potential technology-assisted approaches in standard settings such as use of electronic tools to increase the overall efficiency of the process including data processing and reporting
c) Description of proposed process and approach to be employed in finalizing the standard setting method(s)—to be approved by the PARCC Governing Board and ACCR— in a way that the final method(s) comply with the general technical requirements described in this section of the RFP,
d) Description of potential standard setting methods that best fit the Partnership's needs and requirements,
e) Discussion of how the Partnership should utilize empirical data in setting standards without relying solely on such data,
f) Discussion of how standards will be articulated/aligned across levels within a grade, and across grades,
g) Description of potential methods in estimating decision consistency and accuracy, and variability (standard error) around cut scores.
h) Description of potential quality control procedures to ensure accuracy in data processing

Deliverables for Section V.E.6.

a) Detailed standard setting design and method(s) to be approved by PARCC GB and ACCR (by 6/10/2014) that includes information on training, use of empirical data, round-by-round implementation, panelist feedback, and evaluations and related documentation and presentations
b) Materials to be used for training 16 weeks prior to standard setting for PARCC review and approval

c) Data processing and data reporting plans and tools
d) Evaluation tools, such as surveys, to be used at the standard setting workshop six weeks prior to standard setting for PARCC review and approval
e) Final quality control procedures

Response

Tools for Presenting Materials, Data Processing, and Data Reporting

Pearson proposes a standard-setting process that reflects PARCC's desire to incorporate the use of technology in all aspects of the assessment system. The use of technology-assisted tools in the standard-setting process will provide consistency with the delivery of the PARCC assessments.

Pearson proposes to minimize the logistical difficulties in materials preparation by recommending the use of a 2015 operational form for panelists to experience the test. This will reduce the need to build new test forms for the standard-setting panels.

Pearson proposes to enhance security of materials by delivering the assessments and the ordered item booklets (OIBs) through the test delivery system used for the operational assessments. Therefore, the test security features that are part of operational administrations will also be present for the standard-setting panels.

Pearson proposes minimizing the need for hard-copy materials by delivering the experience the test forms and the OIBs through the test delivery system.

Pearson proposes increasing the overall efficiency of the process including data processing and reporting through the use of individual secure laptops for panelists and implementation of an audience response system which can electronically capture panelists' survey responses and judgments during the standard-setting rounds.

Electronic tools will be used for presenting materials, data processing, and data reporting.

Presenting Materials

Standard-setting materials will be delivered and presented electronically. Each panelist will be provided a secure laptop. Presentations and training materials will be presented through projectors. To enhance panelists' understanding of the tasks required of them during the standard-setting process, clear and concise presentations will be developed in advance of the meetings. This set of materials will then be presented to the panelists by each facilitator. The presentations will be provided to the facilitators as part of facilitator training events to allow for consistency across panels and across events. The presentations for panels will include the review of the PLDs, development of the borderline student descriptors, standard-setting training, feedback data provided after the judgment rounds, and the vertical articulation process.

The test delivery system, which is the proposed tool to deliver the PARCC operational assessments in spring 2015, will be used to allow panelists to experience a sample test on a secure laptop. A tutorial, highlighting key components of the online interface, will be provided to panelists prior to taking the test. The sample test enables panelists to experience the test just like students would. The test delivery system will also be used to provide the practice OIB during the panelists' training for standard setting and to display the OIB during the judgment rounds.

Delivery of the assessments and the OIBs through the test delivery system will significantly reduce the need for hard-copy materials. By eliminating the need for printed test booklets and OIBs, the security of the items included in the experience the test forms and OIBs will be enhanced.

Data Processing

Panelist feedback will be captured through an audience response system. Pearson has successfully implemented the Qwizdom Response System to capture panelists' feedback in real time during standard settings for Virginia, Minnesota, Florida, Georgia, and ReadyPoint. The figure below illustrates one of the Qwizdom Response System electronic tools, or clickers. Each panelist is assigned a clicker that has a distinct ID and relays the panelists' responses in real time to the facilitator's computer. Panelists will be trained to use the Qwizdom clickers during the standard-setting training activities. Pearson's experience with the clickers has demonstrated that they are easy for panelists to use and provide fast and accurate response data. The clickers can capture a variety of feedback from the panelists, including responses to readiness surveys and the page of the OIB bookmarked for each performance level. The facilitator can monitor the panelists' responses in real time thereby allowing early detection of panelists struggling with the standard-setting tasks. All data is captured electronically and easily extracted in a format that can be used as input for statistical analysis. Use of the audience response system will help facilitate faster turnarounds of the feedback data.

Example Clicker for Electronically Capturing Panelist Feedback

Data Reporting

The complexity of the empirical data results and the important role the results play in setting the PARCC performance standards requires clearly articulated presentations of data for a variety of stakeholders (e.g., standard setting panels, pre-policy meeting, etc.). Pearson has extensive experience in synthesizing empirical results. Complex data will be graphically displayed to enhance understanding and to articulate the various findings, focusing on the results most pertinent to the outcomes associated with the PARCC assessments. Several possible displays of empirical results will be shared in Section V.E.6.B.

After each judgment round, panelists will be provided with feedback data. The type of data varies based on the round. Feedback will include individual judgments by the panelists, summary statistics for their table and overall panel recommendations, impact data, and links to external studies. Where applicable, this information will be displayed for the panelists to review and discuss in order to reduce the need for printing materials. The vertical articulation and reasonableness review will be enhanced through visual displays of the performance levels across the grades.

The electronic tools proposed for the standard-setting process have not been developed through federal grants or contracts and will be made available to the Partnership for review upon request.

Quality Control Procedures for Data Processing

Accurate data is an essential component for a successful standard-setting workshop. Use of the Qwizdom tool will allow panelists to transmit their judgments electronically to data analysts. This will eliminate possible data entry errors by data analysts. Although it is possible for the panelists to make data entry errors when using the clicker tool, the tool can be setup to disallow data that do not fit expected values. Additionally, the panelists will be trained to use the tool before making their recommendations. This training includes a practice judgment round where facilitators will verify that the page bookmarked in the OIB matches the page entered into the tool. Panelists will be able to verify the individual feedback they receive on their cut score placements matches where they placed their bookmark. The programs that data analysts use to generate the feedback data and compute the recommended cut scores will be rigorously tested prior to the standard-setting meeting. This includes running test data through the program to verify output, and training the data analysts to perform reasonableness checks on their output. Furthermore, all results from the research studies performed to support standard setting will be verified for accuracy prior to the standard-setting meetings.. Through these quality control procedures, PARCC can be confident in the data from the standard-setting process.

V.E.6.B. Design

The standard-setting process for the PARCC summative assessments will integrate the PARCC's College- and Career-Ready (CCR) Determination Policy, policy-level and subject- and grade-specific Performance Level Descriptors (PLDs), empirical data, and content expert judgment to set five performance levels for each assessment. The CCR Determination Policy describes students' readiness for college and careers for the courses in grade 11. The performance levels for grades 3–11 discuss students' readiness for future grades or in the case of grade 11, students' readiness for college and careers. The use of empirical studies in the standard-setting process to support the policy statements aligns well with PARCC's inclusion of student readiness in the definition of the performance levels.

PARCC has two required summative assessment components: performance based assessment (PBA) and the end-of-year (EOY) assessment. The PBA assessments include both machine-scored and human-scored items. The EOY assessments contain only machine-scored items. The ELA/literacy assessments consist of three item types: Prose Constructed Response (PCR), Evidence-Based Selected Response (EBSR), and Technology Enhanced Constructed Response (TECR). Mathematics assessments consist of three item types: Type I (Tasks assessing concepts, skills, and procedures), Type II (Tasks assessing expressing mathematical reasoning), and Type III (Tasks assessing modeling/applications). Many of the items entail partial credit scoring.

Pearson proposes an Evidence Based Standard Setting process (EBSS; McClarty, Way, Porter, Beimers & Miles, 2013) to integrate empirical data from systematic research and content expert judgment in setting the performance standards for PARCC. EBSS is a process that supports policy claims through systematic research designed to inform the judgments made by content experts. EBSS lends itself well to creating a system of aligned performance standards starting with college and career readiness and linking down from high school to middle school to elementary school. This approach has been used successfully to set performance standards on several assessment programs in recent years, including the American Diploma Project (ADP) and the New York, Virginia, and Texas assessment programs

PARCC is positioned well for the use of an EBSS process through the definitions of the CCR Determination Policy, the policy level PLDs, and subject- and grade-specific PLDs. We believe that the EBSS approach best fits the needs of the Partnership because, if properly planned and executed, it will result in performance standards that not only represent the students' degree of mastery of the assessed curricula based on the Common Core State Standards (CCSS), but can also be used to evaluate their readiness for college and careers. The critical elements of the proposed EBSS approach are described below.

- **Curriculum.** The CCSS, which provide a clear and consistent description of the knowledge and skills students are expected to learn, serve as the underlying basis for several key components of the standard-setting process, including the definitions for each performance level and grade/content area-specific PLDs.
- **Assessment.** Each PARCC assessment has been developed to assess the knowledge and skills described in the CCSS. Each PARCC assessment is based on the requirements described in the anchor standards for the specific grade level and content area and should adhere to the published blueprint and test specifications.
- **Policy Considerations and External Validation.** Results from research studies, which (1) compare performance on the PARCC assessments with scores on other related measures or external assessments and (2) gather post-secondary instructors' judgments on CCR expectations. Stakeholders and experts from across the member states with experience in educational policy and knowledge of the PARCC assessments consider the study results when making recommendations about reasonable ranges for setting each performance standard.
- **Expertise and Knowledge about Students and Subject Matter**. Educators, including classroom teachers and curriculum specialists from K–12 and higher education across the member states, bring content knowledge and classroom experience to the standard-setting process. They play an integral role in developing the PLDs and in recommending the performance standards.

- **Standard Setting.** Within the framework of EBSS, an established standard-setting method known as the bookmark method with external data (Ferrara, Lewis, Mercado, D'Brot, Barth, & Egan, 2011; Phillips, 2012) is used to recommend the cut scores for each PARCC assessment.

Proposed Standard-Setting Process

To implement the EBSS approach for the PARCC assessments, the following seven-step process is proposed.

- Step 1. Define the outcomes of interest and policy goals
- Step 2. Develop research, data collection, and analysis plans
- Step 3. Synthesize the research results
- Step 4. Conduct pre-policy meeting
- Step 5. Conduct standard-setting meeting with panelists
- Step 6. Conduct reasonableness review through post-policy meeting
- Step 7. Continue to gather evidence in support of the standards

A description of each step in the proposed standard-setting process is provided next.

Step 1. Define the Outcomes of Interest and Policy Goals

Before performance standards can be established, the purpose and use for an assessment needs to be clearly articulated for students, parents, educators, and policy makers. The development of the PARCC CCR Determination Policy and the Policy-Level PLDs specify policy claims and general content claims regarding student performance on PARCC assessments which specify the outcomes of interest and the policy goals. PARCC requires five performance levels that delineate the knowledge, skills, and practices that students are able to demonstrate. In addition, PARCC defines the Level 4 performance level as the necessary level for attaining college- and career-ready determination for the grade 11 ELA/literacy assessment and the Algebra II and Integrated Mathematics III assessments. EBSS combines systematic research and content-based expert judgment in a well-defined process to set a coherent system of performance standards that will support the policy and content claims of the PARCC assessments.

Step 2. Develop research, data collection, and analysis plans

Extensive research will be conducted to support the PARCC standard-setting process. The proposed studies should empirically link test scores on PARCC assessments in consecutive grade levels within the same content area (for example, grades 3 and 4 mathematics). Proposed benchmark studies that inform performance on middle school and high school PARCC assessments to related external instruments or measures, such as NAEP, ACT or College Board assessments, international assessments, and applicable state assessments, will also be conducted.

A proposed postsecondary judgment study will gather expert judgment from instructors and professors on PARCC items regarding how well students need to perform to be academically ready for relevant postsecondary courses. In addition, a field-trial of the standard setting will provide valuable feedback for understanding stakeholders' knowledge and use of the empirical data and refining the standard-setting process. Research, data collection, and analysis plans are described in Section V.E.5.

Step 3. Synthesize the Research Results

One of the key elements of the EBSS approach is the incorporation of empirical data in the standard-setting process to help inform content-based judgments made by the standard-setting panels and to provide validity evidence in support of the recommended standards. Pearson proposes using the results of the four standard-setting research studies described in section V.E.5 in the following ways:

a) To help determine *reasonable ranges* prior to the standard-setting meeting

b) To provide *feedback data* to panelists during the standard-setting meeting; and

c) To evaluate the *reasonableness* of the recommended cut scores across PARCC assessments during vertical articulation at the end of the standard-setting meetings

Reasonable Ranges

The reasonable ranges provide boundaries within which standard-setting panelists can recommend the placement of cut scores that will define each of the PARCC performance standards. The empirical results from studies 1-4 (see section V.E.5) will help to inform reasonable ranges for the Level 4 cut score at key grade levels. The underlying vertical scale and student performance on the field test and operational assessments can help inform reasonable ranges for the remaining cut scores and reasonable ranges for other grade levels. These ranges can be evaluated across grade levels within a content area to verify that they align well. The alignment can be evaluated using impact data at each grade, vertical scale scores, and linkages to external measures. Reasonable ranges can also be compared across subjects within a grade level using impact data and linkages to external measures.

Through the results from the benchmark study (Study 1) for example, performance differences in PARCC states on NAEP mathematics and ELA could be used to check the reasonableness of impact data for PARCC mathematics and ELA based on the reasonable ranges developed for the two tests. If the reasonable ranges for PARCC mathematics appear more rigorous than the PARCC ELA reasonable ranges in terms of impact data, but NAEP would suggest that PARCC states tend to perform better on mathematics than on ELA, this might suggest that the reasonable ranges are not well aligned across content areas. Of course, consideration would need to be given to other factors, for example, differences in content coverage between the two assessments.

Feedback Data

Results from the empirical studies will also be used during the standard-setting workshop as part of the feedback data provided to panelists for one or more of the rounds of judgments. Such feedback could help inform the relationship of the panel's recommended cut score to external tests or criteria. The proposed feedback for each round of judgment is discussed in Step 4. The following figures provide illustrations of how the empirical studies can be synthesized for feedback data.

For example, after each round of judgment, the panelists could be provided with the projected SAT score and/or the likelihood of meeting the corresponding ACT college readiness benchmark, based on the panel's recommended cut score for PARCC Level 4. The following figure illustrates what could be shown to the standard-setting panelists.

PARCC Algebra II—Round 1	
Based on the currently recommended cut score, a borderline Level 4 student's:	
Projected SAT Mathematics Score is	513
Likelihood of Meeting the ACT Mathematics College Readiness Benchmark (of 22) is	63%

Example Standard-Setting Panel Feedback Data

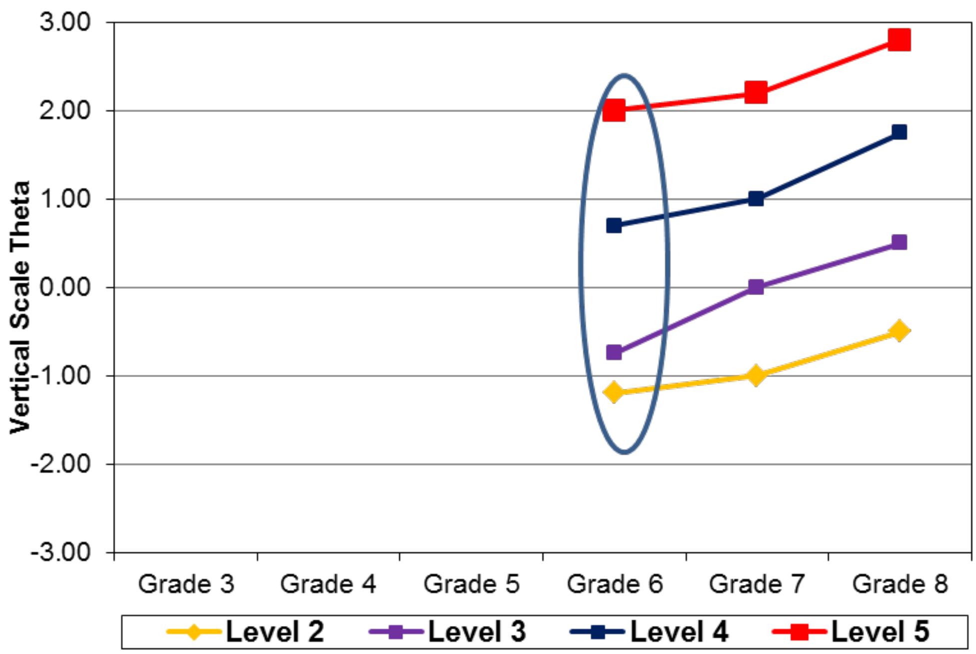
Study results can also provide information to help panelists' align their recommendations with cut scores at upper grade levels in the same content area. For example, when panelists recommend cut scores for the grade 8 assessments, feedback relative to the high school recommended cut scores can be provided to the grade 8 panel. In addition, when panels are setting cut scores for two grade levels, the feedback for the second assessment can include student performance based on the panel's recommended cut scores for the first assessment. The following figure provides an example of this type of feedback data where the likelihood of students meeting the corresponding cut score in the next grade-level is provided based on the current recommended cut for a lower grade-level. This feedback presents empirical data for the panelists to consider as they make content judgments.

PARCC Grade 5 Reading—Round 2	
Cut Score	**Minimum likelihood of meeting the corresponding cut score in Grade 6 Reading**
Level 2	73%
Level 3	69%

Level 4	81%
Level 5	54%

Example Standard-Setting Panel Feedback Data for the Grade 5 & 6 Panel

The underlying vertical scale can also be used to show the alignment of the panel's current recommended cut score compared to upper grade-level recommended cut scores. The following figure provides an illustration for a grade 6 panel. In this example, the recommended cut scores for the grades 7 and 8 tests are shown in relation to the current grade 6 recommended cut scores.

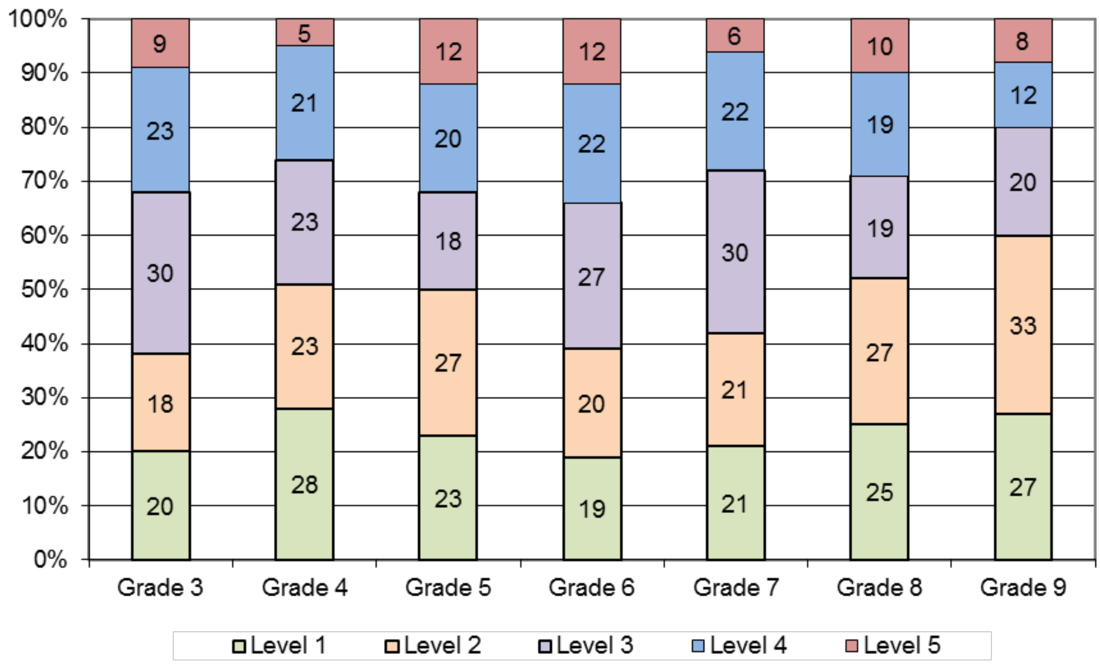


Example Vertical Scale Feedback Data for the Grade 5 & 6 Panel

Panelists can use these types of feedback data along with additional information, such as impact data, to evaluate the reasonableness of their cut score recommendations and make adjustments accordingly.

Reasonableness Review

During the vertical articulation at the end of the standard-setting workshop, panelists will review the PLDs for all grades within a subject to gain perspective on the rigor associated within and across grades. They will discuss what their expectations are for cut scores and impact data given their understanding of the rigor. They will review the third round recommended cut scores from all the grades as well as the associated impact data. Impact data are the percent of students that would be in each performance level if the round three cut scores were adopted. Panelists will be asked to discuss similarities and differences across grade levels in the cut scores relative to the rigor of the expectations as outlined in the PLDs and make individual recommendations for adjustments, if warranted, to maintain consistency of rigor within a subject area. The figure below provides an example for displaying the percent of students classified into each performance level following the round 3 recommendations during vertical articulation. This impact data based on the recommended performance cut scores illustrates the five performance levels from grades 3-9.



Example Reasonableness Review Slide for Recommended Performance Standards

Step 4. Conduct Pre-Policy Meeting

A pre-policy meeting comprised of stakeholders such as the PARCC Governing Board and Advisory Committee on College Readiness (ACCR), Technical Advisory Committee (TAC) members, and Partnership representatives can be convened to evaluate study results and weigh in on reasonable ranges. The stakeholders would consider policy implications of the performance standards along with the empirical study results to make recommendations for reasonable cut score ranges on the PARCC assessments.

A “briefing book” approach (Haertel, Beimers & Miles, 2012) can be implemented in which stakeholders are provided a binder composed of research study results highlighting the potential reasonable ranges along with related impact data and other relevant information to help inform their evaluation and recommendations for the reasonableness ranges to present at the standard-setting event. The briefing book would summarize the research results from studies 1-3. The research results may be displayed in table format as shown in the following figure, which lists the PARCC Raw Score, Theta, impact data, benchmark study results and post-secondary educator judgments studies. The briefing-book will summarize the quality of each of the studies in terms of sample size, assessment overlap, statistical relationship, representativeness of the sample relative to the population, and motivation.

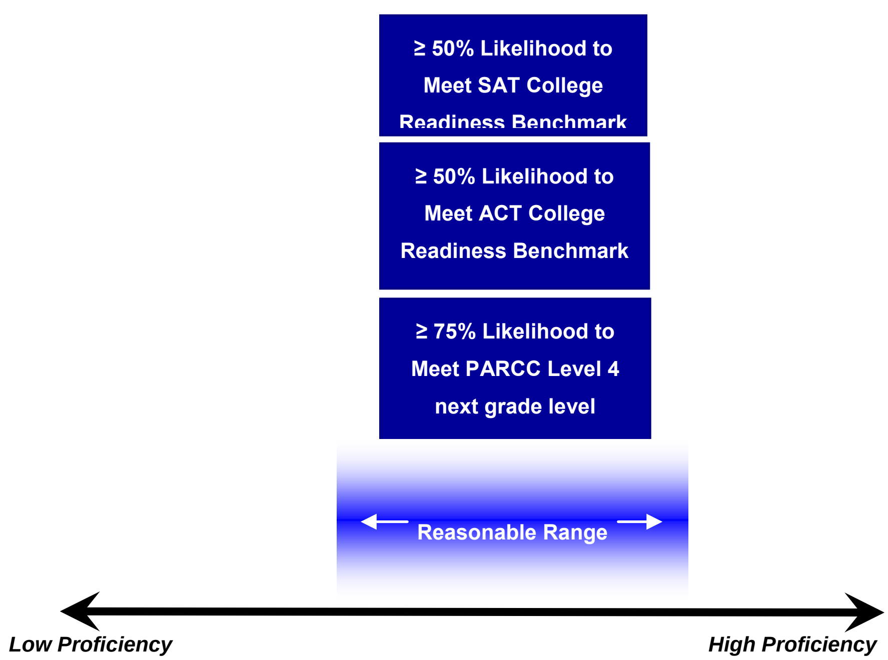
PARCC Raw Score	PARCC Theta	Impact Data	Benchmark Studies Impact Data	Judgment Studies Cut Scores
0	-4.091	100		
⋮	⋮	⋮		
70	0.478	80		
71	0.514	79		
72	0.523	75		
73	0.542	71		
74	0.633	66		
75	0.794	65		
76	0.898	62		
77	0.898	58		
78	1.024	55		
79	1.128	54		
80	1.204	53		
81	1.454	51		2YT Composition
82	1.756	50		
83	1.808	47		2YT Technical
84	2.068	45		2YT Literature
85	2.309	41		
⋮	⋮	⋮		4YO Composition
110	2.432	31		
111	2.812	30	NAEP Proficient (30)	4YS Composition
112	2.820	27		
113	2.836	24		
114	2.952	20		
115	2.973	18		
116	3.023	17		
117	3.159	15		4YO & 4YS Literature
118	3.404	14		
⋮	⋮	⋮		
130	3.719	10		
131	3.728	7		
132	3.733	5		
133	4.633	3		
⋮	4.678	2		
150	4.997	1		

It is often helpful to develop initial guidelines for establishing the reasonable ranges prior to the pre-policy meeting which panelists refine during the discussion.

The guidelines should be grounded in the performance level definitions for the cut scores and informed by the list of available research studies for each assessment. The guidelines could, for example, include rules such as:

- For the PARCC Level 4 standard, these study results should be within the reasonable range:
  - At least 50 percent likelihood of meeting the SAT college readiness benchmark in the related content area;
  - At least 50 percent likelihood of meeting the ACT college readiness benchmark in the related content area;
  - At least 75 percent likelihood of meeting the PARCC Level 4 standard at the next grade level in the same content area;
  - At least 75 percent likelihood of earning college credit by attaining at least a grade of C or higher in the next relevant postsecondary course for the grade 11 ELA/Literacy, Algebra II, and Integrated Mathematics III assessments.

The following figure provides an illustration of how various validity studies could be used to determine reasonable ranges for a PARCC assessment. Factors such as data quality, representativeness of study sample, and overlap in assessed curriculum between the empirically-linked assessments could be used to help determine which research study results are given higher priority in determining the final reasonable ranges.



Visual Representation of Example Guidelines for Reasonable Ranges

Once the reasonable ranges for a PARCC assessment have been established, they can be used in the OIB construction process. The OIB will represent the content of a PARCC assessment but will mostly consist of items whose difficulty values fall within the reasonable ranges. OIB development is discussed in Step 5.

Step 5. Conduct Standard-Setting Meeting with Panelists

Panels consisting of K–12 educators, school administrators, higher education faculty, parents, and other stakeholders can use the PLDs, reasonable ranges, and feedback data based on results from the empirical studies to recommend cut scores for each PARCC assessment. In order to have a comprehensive and aligned assessment system the performance standards for the high school assessments will be recommended first. Then the performance standards for the middle school and elementary school assessments will be established starting with grade 8 and working down to grade 3. This organization of the standard-setting panels will allow for the recommended standards of the upper grade-level panels to serve as feedback for the lower grade-level panels. The following figure lists the schedule and organization of the assessments within the standard-setting panels.

PARCC Standard-Setting Schedule				
Panel	**Start Date**	**End Date**	**Subject/Grade**	
			Mathematics Panels	**ELA/Literacy Panels**
Grades 3–6	8/24/2015	8/28/2015	Grades 3 and 4 Grades 5 and 6	Grades 3 and 4 Grades 5 and 6
Grades 7–8	8/17/2015	8/21/2015	Grades 7 and 8	Grades 7 and 8
High School	7/27/2015	7/31/2015	Algebra I and Integrated 1 Geometry and Integrated 2 Algebra II and Integrated 3	Grade 9 Grade 10 Grade 11

Panelist Selection

Panelists from across the PARCC partnering states will be convened to recommend cut scores on all the PARCC assessments. Each panel will be asked to recommend four cut scores resulting in five performance levels on each PARCC assessment. In making their recommendations, the panelists will consider the following types of information:

- Assessed curriculum in CCSS
- Items on the PARCC assessments
- Performance level definitions and specific PLDs for each assessment
- Reasonable ranges for each performance standard
- Selected results from the standard-setting research studies

When selecting standard-setting panelists, Pearson recommends placing an emphasis on content knowledge and classroom experience. The judgments and cut-score recommendations made by the panelists, however, will also be informed by empirical studies, both through the reasonable ranges and as feedback provided after each round of judgment. The composition of the standard-setting panels will follow the requirements outlined in Section V.E.4.

Each panel should also be representative in terms of demographics (e.g., gender and ethnicity) and student populations (e.g. special education, English language learners). Panels should include individuals from different positions with varying years of experience in education, and from different types of campuses and districts (e.g., large vs. small, urban vs. rural etc.) In addition to educators from the PARCC states, Pearson will work with the PARCC state testing directors to recruit panelists from other stakeholder communities. The grades 3-8 panels will be comprised of representatives of grade-level teachers, above grade-level teachers, parents, and business community representatives as appropriate. The higher grade panels will be comprised of representatives of grade-level teachers, technical or vocational educators, higher education faculty or staff, non-teacher educators (e.g., curriculum specialist), and parents or other stakeholders (business representatives or chamber of commerce members) as appropriate. To help enhance the continuity of the standard-setting process, PARCC could consider inviting educators who also served on the PLD and pre-policy meetings as well as post-secondary representatives who participated in the research studies so that they can share their experiences. Specific details about the recruitment of panelists were previously presented in Section V.E.4.

The standard-setting process is complex and requires a great deal of integration of information on the part of the panelists. Pearson facilitators are trained to gauge panelists' understanding and have experience in redirecting discussions as needed to help panelists stay on track and work effectively and efficiently. When panelists have concerns or misunderstandings, the facilitators will work to resolve issues before moving on.
Despite best efforts to recruit knowledgeable and helpful standard-setting participants, occasionally a panelist will be unable or unwilling to carry out the standard-setting process. When this happens, facilitators will notify the Standard-Setting Lead who will work with PARCC leadership to find a solution. Steps will be taken to maintain a coherent execution of the workshops.

Standard-Setting Workshop Agenda

The following figure provides a sample agenda for the PARCC standard-setting workshop. Pearson will collaborate with PARCC in terms of the timing of the agenda.

Standard-Setting Agenda	
General Session	▪ Overview and Purpose of the PARCC Assessments ▪ Overview and Purpose of Standard Setting
Breakout Session	▪ Online Tutorial for Experiencing a PARCC Test ▪ Experience a Test (PBA assessment and EOY assessment)* ▪ Review of Performance Level Descriptors* ▪ Borderline Student Descriptors Development* ▪ Standard-Setting Training, Clicker Training, and Practice OIB ▪ Round 1 Judgments and Feedback* ▪ Round 2 Judgments and Feedback* ▪ Round 3 Judgments and Feedback* ▪ Vertical Articulation ▪ Standard-Setting Evaluation

***These tasks will be repeated for each assessment for which the panel is recommending standards.**

The general session serves to welcome the various stakeholders and panelists. An introduction to the PARCC assessments and description of the standard-setting purposes provides valuable information so that all panelists begin the process with a common understanding of the PARCC assessments and the panelists' role in the critical task of setting performance standards. Panelists will learn about the innovative items administered in the performance based assessment (PBA) and an end-of-year (EOY) assessment. The development of the six priority purposes of the PARCC assessments and the proposed interpretations and uses of the PARCC assessment results will be provided to the panelists through a summary of steps 1-3 in the EBSS process. An overview of EBSS will highlight the empirical research studies that were used in the development of the reasonable ranges.

The breakout session divides the panelists into content specific rooms with approximately 20 panelists each. Panelists will be assigned to tables of four to five panelists such that each table represents a diverse group. The breakout session is where the majority of the panelists' time will be spent. The following sections provide a description of the topics outlined in the breakout session of the sample agenda.

Online Tutorial

The PARCC assessments are computer-delivered and include technology-enhanced items (TEIs). The online tutorial will orient the panelists to the tools available when taking the assessment and the type of items included in the PARCC assessments. The online tutorial will be the same tutorial provided to students during the operational administration of the PARCC assessments so that panelists can become familiarized with the user interface and types of item interactions they will encounter.

Experience a Test

The summative components for each PARCC assessment include a PBA and EOY assessment. The performance standards will be established for the combined performance on PBA and EOY. Therefore, the panelists will experience each assessment on a computer in the same way students experience it.

Review of the Performance Level Descriptors

PLDs are statements that articulate the specific knowledge and skills students typically demonstrate at each performance level of an assessment given for a specific grade level and content area. The PLDs developed for PARCC assessments provide a snapshot of students' academic characteristics and reflect the breadth and depth of the content, skills, cognitive demand, and performance requirements evident in the CCSS. The PLDs were developed as an aligned system, describing a reasonable progression of skills within each content area. For the standard-setting process, the PLDs will be revised as discussed in Section V.E.2. The PARCC grade- and content-specific PLDs describe student performance for the middle of the performance level.

Borderline Student Descriptors Development

After reviewing the PLDs, panelists will be asked to think about the group of students who just barely reach a performance level. These are the "borderline" students—defined as those students who have the minimum amount of knowledge necessary to be in a performance level. Since PARCC assessments have five performance levels, panelists will be asked to define borderline descriptors for four groups of students:

- Level 2 borderline students (cut score between Level 1 and Level 2)
- Level 3 borderline students (cut score between Level 2 and Level 3)
- Level 4 borderline students (cut score between Level 3 and Level 4)
- Level 5 borderline students (cut score between Level 4 and Level 5)

Panelists will work in their table groups to develop descriptors that characterize what a borderline student should know and be able to do. Whereas the PLDs describe students in the middle of a performance level, the borderline descriptors focus on students with just enough knowledge to get them into a performance level. The panel as a whole will discuss the table-level descriptors in order to develop a master set of borderline descriptors. Panelists will use the borderline descriptors while making their judgments.
The following figure illustrates the relationship between borderline students where performance standards are recommended and the typical students as defined by the PLDs.

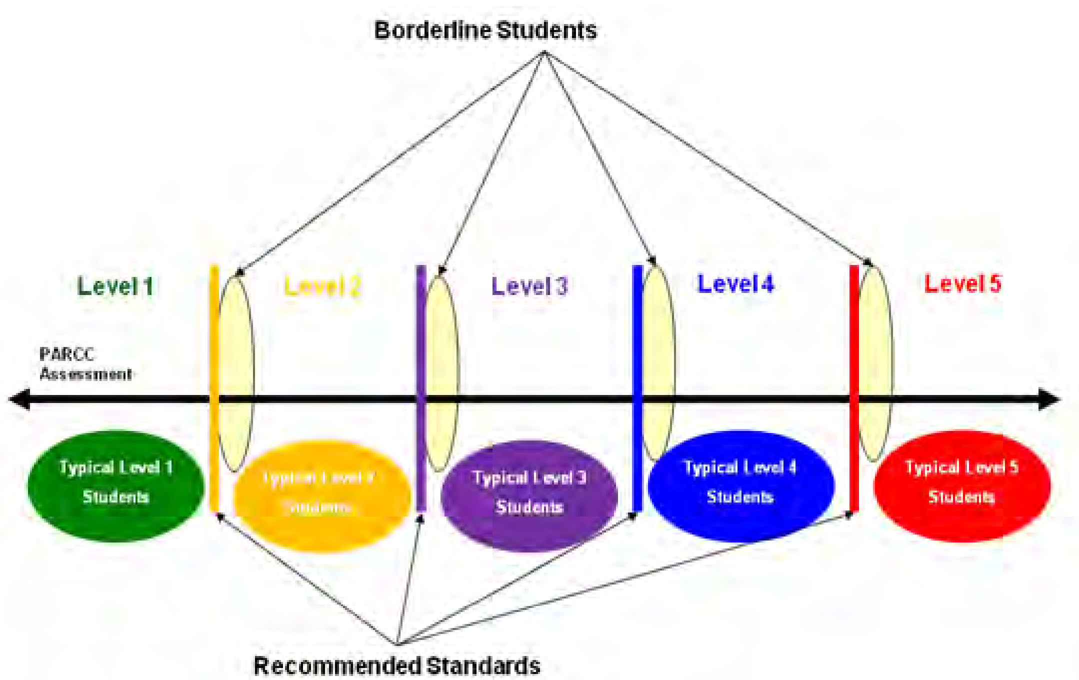


Illustration of the Borderline Students

Standard-Setting Training

The panelists are trained on and follow an established standard-setting method known as the bookmark method with external data (Ferrara et al., 2011; Phillips, 2012) to make their cut score recommendations. One key component of this method is the OIB, which contains items from the PARCC assessments of interest, ordered from least difficult to most difficult in terms of empirical item difficulty. OIBs used by the standard-setting panels are created to reflect the range of difficulty represented by the content standards and field test data and support panelists' evaluations of the reasonable ranges recommended by the pre-policy panel.

Development of the OIB

The PBA assessments include both machine-scored and human-scored items. The EOY assessments contain only machine-scored items. The PBA and EOY assessments will be calibrated on the same theta metric and the final reported score will be based on student responses to both the PBA and EOY assessment.

The innovative item types (PCR, EBSR, and TECR) on the PBA and EOY assessments are critical considerations when developing the OIB. Pearson proposes a combined OIB with items from the PBA and EOY assessments. The items will be ordered by difficulty and the items with partial credit scoring will be represented multiple times in the OIB based on the number of possible score points. The OIB may be augmented with additional items in order to have a variety of items within the draft reasonable ranges for the performance standards.

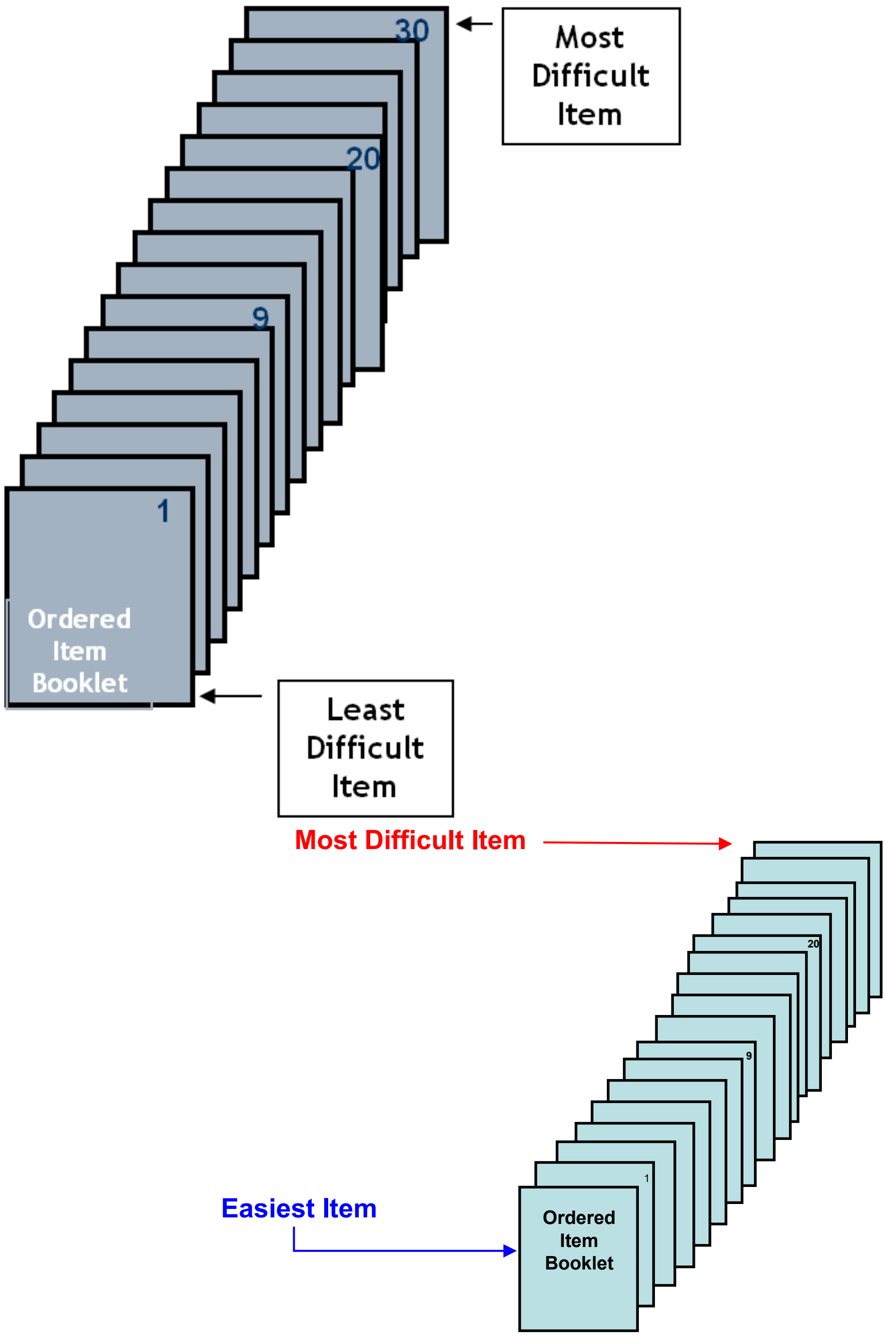
Given the key role of the OIB in this process, each OIB should be carefully constructed to give committee members the most information about the types of items falling within the reasonable ranges recommended by the pre-policy panel. This may require more items that are representative of the difficulty within the reasonable ranges—and therefore relatively fewer items with difficulty values that fall outside the reasonable range—to be included in the OIB. Doing so allows panelists to make finer distinctions between items within the area that the pre-policy panel recommended the standards be set.

Based on the reasonable ranges recommended by the pre-policy panel, each OIB is evaluated to make sure that the regions of the IRT-based scale range of the PARCC assessment that correspond to the reasonable ranges are represented by the items in the OIB. Areas of the OIB that do not have item representation are identified as gaps. Areas of the OIB with an overrepresentation of items are identified as clusters. Additional items are used to fill in gaps in the OIB; items are removed to eliminate clusters.

OIB Implementation

During the standard setting, panelists review the items in the OIBs and place a bookmark following the item that they determined best represents the minimum expected performance for each performance level. By suggesting that panelists place a bookmark within the reasonable range for each performance standard, it helps verify that the resulting recommended cut scores are reasonable and result in an aligned system based on the research studies.

Before the panelists recommend standards, they should receive training on the bookmark procedure. The bookmark procedure requires panelists to review a set of items and decide which of them are likely to be answered correctly by borderline students. The set of items in each OIB are ordered from least difficult to most difficult (see the following figure). As the items become progressively more difficult, panelists decide item-by-item whether a borderline student within the given performance level would be likely to respond correctly.



Arrangement of Items in an Ordered Item Booklet (OIB)

Practice OIB

After the panelists are familiarized with the bookmarking method and the OIB, they will also be trained to use the Qwizdom clicker tool to submit their recommendations. Next, they will participate in a practice activity to make sure they understand the bookmark standard-setting method. In the practice activity, panelists are provided with a practice OIB. The practice OIB consists of 8-10 items ordered by difficulty. The panelists will consider one of the borderline groups (e.g., the Level 2 borderline group). They will read each item and determine whether a borderline student would respond to the item correctly. Panelists will select an item in the practice OIB to place their bookmark representing where borderline students will no longer be able to answer items correctly, and submit the page number of the last item that the borderline student would respond correctly to with the clicker tool. Once panelists complete this process, the facilitator and the panelists will discuss where the bookmark was placed. The practice OIB provides panelists with an opportunity to practice the method of setting their bookmark and using the audience response system or clicker, as well as a brief introduction to how discussions will be structured following each round of recommendations.

Round Judgments

Prior to each judgment round, panelists will be asked a series of questions to verify that they understand their task and are ready to begin. Responses to these questions will be submitted using the clicker tool. Facilitators will review the answers, and clarify any remaining questions for the panelists.

Once panelists are ready to begin making judgments using the OIB, they will begin with the lowest borderline student group, borderline Level 2, and consider the likelihood of the borderline student responding to each item within the reasonable range correctly. Once panelists select the item for placing their bookmark for the borderline student, the panelists will focus on the next borderline group and follow the same process until the panelists have identified bookmarks for each borderline group. Finally, they will use the clickers to record their bookmarked page numbers for each of the four performance level cuts.

Round Feedback

After each judgment round, feedback data is provided to panelists. Feedback data can include results from Studies 1-4 where available, impact data (that is, the percentage of examinees in each performance level), and information about how the cuts line up with cut score recommendations from higher-grade levels. Certain types of feedback data might be provided only after certain rounds. For example, impact data is often held until after Round 2 so that committee members are not overly swayed by pass rates and continue to base their recommendations on a variety of considerations, most importantly, the borderline descriptors. The following figure provides an example of the types of feedback that could be shown to the panelists after each round.

Sample Feedback for Each Round
Round 1 Feedback
▪ The panelists' individual Round 1 cut-score recommendations (bookmarked pages) for each performance level ▪ Table-level Round 1 cut-score recommendations—the minimum, maximum, mean, and median bookmarked pages for each performance level ▪ Panel-level Round 1 cut-score recommendations—the minimum, maximum, mean, and median bookmarked pages for each performance level ▪ The percentage of students answering each item in the OIB correctly (p-values)
Round 2 Feedback
▪ The panelists' individual Round 2 cut-score recommendations (bookmarked pages) for each performance level ▪ Table-level Round 2 cut-score recommendations—the minimum, maximum, mean, and median bookmarked pages for each performance level ▪ Panel-level Round 2 cut-score recommendations—the minimum, maximum, mean, and median bookmarked pages for each performance level ▪ Data showing projections from the panel-level Round 2 cut-score recommendations to the next grade-level assessment in the content area (all cuts) and, where possible, to external tests or measures (for Level 4 cut only) ▪ Impact data (percentage of student in each performance level) for the assessment based on the panel's Round 2 cut-score recommendations. Impact data can be given for the total group of students and by disaggregated student groups (such as gender, ethnicity and special populations) ▪ Impact data from applicable national or international tests (such as NAEP, TIMSS, PIRLS, SBAC, ASVAB or other state assessments) that can be used for comparison purposes with the impact data for the PARCC assessment
Round 3 Feedback
▪ The panelists' individual Round 3 cut-score recommendations (bookmarked pages) for each performance level ▪ Panel-level Round 3 cut-score recommendations ▪ Data showing projections from the panel-level Round 3 cut-score recommendations to the next grade-level assessment in the content area (all cuts) and, where possible, to external tests or measures (for Level 4 cut only) ▪ Impact data (percentage of student in each performance level), for total group and by disaggregated student groups, for the assessment based on the panel's Round 3 cut-score recommendations ▪ Impact data from applicable national or international tests (for comparison purposes)

Vertical Articulation

As a final step in the standard-setting meeting, often the panelists, or representatives from each panel, meet to review the performance standards across assessments. During this articulation session, panelists will be provided a chance to compare their final recommendations to the final recommendations in other grades. Feedback data including vertical scale information, impact data, and study information is provided based on the recommendations from the final judgment round. At this time, if the panelists notice a misalignment, they have an opportunity to discuss and possibly adjust their recommendations. Recommended changes made during the vertical articulation must have a well-articulated rationale. For example, if vertical scale data suggests a change, panelists should consider the OIB and borderline descriptors to see if the change can be supported by content considerations.

For the high school PARCC assessments, panelists who participated in recommending ELA standards would receive feedback data across all three high school ELA assessments. Likewise, panelists who participated in recommending mathematics standards would receive feedback data across all six high school mathematics assessments. For the grades 7-8 panels, participants would be provided feedback for ELA or mathematics. The approved high school cut scores would also be available in time for the grades 7-8 standard-setting workshop, and could be provided for comparison. Likewise, during the grades 3-6 meetings, panelists would be provided feedback for the PARCC assessments they worked on, as well as information about how their recommended cut scores line up with the recommended cut scores in 7-8 and high school.

Alignment of the High School Mathematics Courses

The panels for the high school mathematics courses will recommend performance standards for the traditional and integrated high school mathematics end-of-course assessments. For example, the same panel will recommend Algebra I and Integrated Mathematics I standards. The panelists will use the same EBSS process to recommend cut scores on both assessments. They will base their recommendations on the same considerations of empirical data and content considerations including the PLDs and borderline student descriptors. Use of the same process and understanding of borderline students should promote highly comparable standards on the traditional and integrated mathematics assessments. In addition, impact data, vertical scaling data (if available), and empirical data from studies 1-4 can be used during the articulation to verify the comparability and coherency of the mathematics standards.

Pearson proposes using the results of the comparability study that will be conducted using field-test data to determine the relationship between the IRT scales for the traditional and integrated mathematics assessments. Pearson proposes conducting similar analyses using the operational data from the 2015 administration to augment the prior study results. Based on the study results, the recommended performance standards will be displayed relative to the traditional and integrated mathematics assessments.

To achieve meaningful comparability between cut scores for the traditional and integrated mathematics assessments, panelists will consider the recommended cut scores, empirical research studies, and the PLDs to discuss the alignment of the mathematics cut scores.

During the vertical articulation, panelists can be provided feedback data that relates the panel's recommend Level 4 cut scores for the traditional and integrated mathematics to the next grade level. The following figure is an example of how the feedback data may be presented. Panelists can discuss the reasonableness of the feedback data given the Level 4 borderline descriptors for the traditional and integrated mathematics courses.

Example Feedback Data for Traditional and Integrated Mathematics		
Performance Standard	**Level 4 Algebra I**	**Level 4 Integrated Mathematics 1**
Probability of reaching the Level 4 cut score in Geometry	75	67
Probability of reaching the Level 4 cut score in Integrated Mathematics 2	85	75

Standard-Setting Evaluation

One of the final tasks for the panelists will be completing surveys evaluating the standard-setting process including their understanding of the process, their use of empirical research and content judgments in making recommended cut scores, their review of the PLDs, and their overall impression of the recommended standards from their panel. Pearson can work with PARCC to design an electronic standard-setting evaluation survey that panelists fill out on their individual laptop computers.

Step 6. Conduct Reasonableness Review via a Post-Policy Meeting

A post-policy panel will be convened to review the cut scores recommended by the standard-setting panels across grade levels and content areas to evaluate the reasonableness of the performance standards as a system and make policy-based adjustments as appropriate. Detailed information about the post-policy panel is provided in Section V.E.7.

Step 7. Continue to gather evidence in support of the standards

After the PARCC performance standards are finalized, additional evidence can be gathered in support of the reliability and validity of the cut scores. In particular, the longitudinal study discussed in Section V.E.5.E. will examine the external validity of the cut scores for PARCC CCR performance levels by comparing student performance in college to their previous scores on PARCC assessments.

As required input to the Section V.E.7., two types of error can be evaluated in relation to cut scores on PARCC assessments: measurement error and standard-setting error.

Measurement Error

Measurement error can be quantified using measures such as the conditional standard error of measurement, decision consistency, and decision accuracy. The conditional standard error of measurement provides an estimate of the amount of variability at each score point in the score scale resulting from factors other than what the assessment is designed to measure. When using an IRT model, typically measurement error is greatest at low and high scores points where few students score, because of the lack of data with which to estimate ability. In the middle of the score scale, where many students score, the conditional standard error is typically the smallest. Less measurement error results in a higher degree of confidence in the scores students are assigned, that is, more reliable scores. Once operational PARCC forms are created, pre-equating, using field-test statistics might be used to generate a raw score to theta table and the conditional standard error associated with each raw score.

With the five performance categories that will be established on PARCC, it is also important to see how consistently and accurately students are categorized. Measurement error can result in some students being misclassified into a lower or higher category than the category that truly reflects their level of academic achievement. It is important to evaluate decision consistency and accuracy to confirm that the PARCC assessments reliably and accurately categorize students such that the decisions made based on test scores are appropriate. A variety of indices are available to quantify decision consistency and accuracy (see for example Kim, Choi, & Um, 2006; Lee, Hanson, & Brennan, 2000). Methods can be selected that appropriately estimate decision consistency and accuracy given the number of PARCC performance categories and the measurement model used. Field-test statistics for the operational PARCC forms can be used to provide estimates of decision consistency and accuracy.

Standard-Setting Error

Another source of error that impacts cut scores is the variability involved in establishing the cut scores during the standard-setting meetings. Measurement error is often described as the variability in students' scores if they could be administered an assessment on multiple occasions but not remember taking the test previously. Similarly, if a standard-setting committee could go through the standard-setting process on multiple occasions, but not remember setting standards on the assessment previously, there might be variability in the chosen cut scores. In the case of the PARCC standard-setting meetings, very few representatives can be included from each state in order to maintain approximately 20 panelists per panel. If different samples of representatives were chosen, different cut score recommendations might be obtained.

By holding two or more standard-setting meetings for each grade and subject, cut score recommendations could be compared across committees. However, replicating the standard-setting process is not usually logistically or financially feasible. Additionally, the two sets of cut scores might not be true independent replications of the standard-setting process. For example, a truly independent replication would require a different facilitator and a completely different set of panelists.

Creating separate table groups within each committee can be used to provide estimates of consistency in cut score recommendations within a committee. For example, PARCC intends to have approximately 20 participants per committee. Participants could be divided into 4-5 table groups. The median cut scores after the final judgment round for each table could be used to evaluate variability in cut score recommendations. Another way to estimate variability would be to calculate the variability in cut score recommendations across all 20 panelist recommendations after the final judgment round. However, the recommendations within a committee cannot be considered independent because participants work as a group to develop an understanding of the performance levels.

Despite the challenges of estimating variability in the standard-setting process, there are several criteria that can be used to promote accuracy and reliability in the standard-setting process. Cizek, Bunch, and Koons (2004) categorize these criteria into the following three types:

1. **Procedural.** This includes establishing well-articulated purposes and processes for the standard setting prior to workshops; making sure the standard-setting process is feasible, credible, and interpretable to all parties involved; implementing procedures with fidelity during the standard-setting meetings; providing feedback to committee members throughout the standard-setting process such that they have confidence in the process and in the results; and documenting the standard-setting method and results.
2. **Internal.** This includes the variability that would be expected if the standard-setting process could be repeated with multiple committees; the variability within a committee member; and the variability across committee members. Variability within and across committee members cannot be evaluated in terms of "random replications" because the committee members are not blank slates after each round of judgment (nor would we want them to be), but recommendations can be compared within and between judgment rounds.
3. **External.** This includes the variability that could be expected between the standard-setting method chosen and other appropriate standard-setting methods; comparing standard-setting results to other important external measures including course grades and measures of similar constructs; and evaluating the reasonableness of cut scores including pass rates overall and by subgroup.

As discussed above in Steps 1-7, there is a well-established process for setting standards within the EBSS framework. Pearson has extensive standard-setting experience and will work with PARCC to implement these steps with fidelity. Additionally, Pearson can use information from the standard-setting judgment rounds to provide estimates of variability within and between committee members. Information collected in Studies 1-5 can be used to provide external validity evidence indicating that the PARCC standard-setting process resulted in valid performance standards.

In summary, using the EBSS process proposed by Pearson provides a framework for establishing reliable and accurate PARCC performance standards.

References

Cizek, G. J., Bunch, M. B., & and Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice, 23*(4), 31-50.

Ferrara, S., Lewis, D., Mercado, R., D'Brot, J., Barth, J., & Egan, K. (2011, April). *A method for setting benchmarked performance standards: Workshop procedures, panelist judgments, and empirical results.* Paper presented at the annual meetings of the National Council on Measurement in Education. New Orleans, LA.

Haertel, E. H., Beimers, J. N., & Miles, J. A. (2012). *The briefing book method*. In G. J. Cizek (Ed.), *Setting performance standards* (2nd ed.). New York: Routledge.

Kim, D.-I., Choi, S. W., & Um, K. R. (2006). *A comparison of methods for estimating classification consistency.* Paper presented at the annual meeting of the National Council on Educational Measurement, San Francisco, CA.

Lee, W.-C., Hanson, B. A., & Brennan, R. L. (2000). Procedures for computing classification consistency and accuracy indices with multiple categories. *ACT Research Report Series.* Retrieved December 4, 2013 from http://www.act.org/research/researchers/reports/pdf/ACT_RR2000-10.pdf

O'Malley, K., Keng, L., & Miles, J. (2012). Using validity evidence to set performance standards. In G. J. Cizek (Ed.), *Setting performance standards* (2nd ed.). New York: Routledge.

Phillips, G. W. (2012). The Benchmark Method of standard setting. In G. J. Cizek (Ed.), *Setting performance standards* (2nd ed.). New York: Routledge.

V.E.7. Review of Cut Scores

Requirement

Response Requirements for Section V.E.7.
a) Description of the approach and procedures to complete all the responsibilities/tasks specified in Section V.E.7
b) Description of proposed method and examples of the materials for presenting the cut scores to the PARCC Governing Board and to ACCR

Deliverables for Section V.E.7.

a) Detailed design and methodology for the policy makers' review of cut scores
b) Initial cut scores, impact data, decision consistency and accuracy, and variability (standard error) around cut scores and other relevant information to be presented to PARCC Governing Board and ACCR
    i. The initial cut scores and impact data for all assessments shall be presented to the PARCC states as soon as they are available for their review (before they are presented to the PARCC Governing Board and ACCR approval)
c) Summary of the outcomes of the Governing Board and ACCR review sessions
d) Tools and materials that could be used for cut score adjustments six weeks prior to the Governing Board and ACCR review
e) Final cut scores, impact data, decision consistency and accuracy, and variability (standard error) around cut scores and other relevant information in format to be approved by PARCC

Response

Policy Makers' Review of Cut Scores

Post-policy panels will be convened to review the cut scores recommended by the standard-setting panels across grade levels and content areas to evaluate the reasonableness of the performance standards as a system and make policy-based adjustments as appropriate.

PARCC post-policy panels will include state representatives, the PARCC governing board, and the ACCR. Because the high school standards are set before the 3-8 standards, there will be post-policy panel reviews to evaluate the recommended high school standards. Later, additional post-policy panel reviews will be conducted for the 3-8 standards. There are four types of post-policy panel reviews that will take place in the following order:

1. PARCC state review of the recommended high school standards
2. PARCC governing board and ACCR review of the recommended high school standards
3. PARCC state review of the recommended 3-8 standards
4. PARCC governing board review of the recommended 3-8 standards

These four types of post-policy panel reviews are described in the pages that follow.

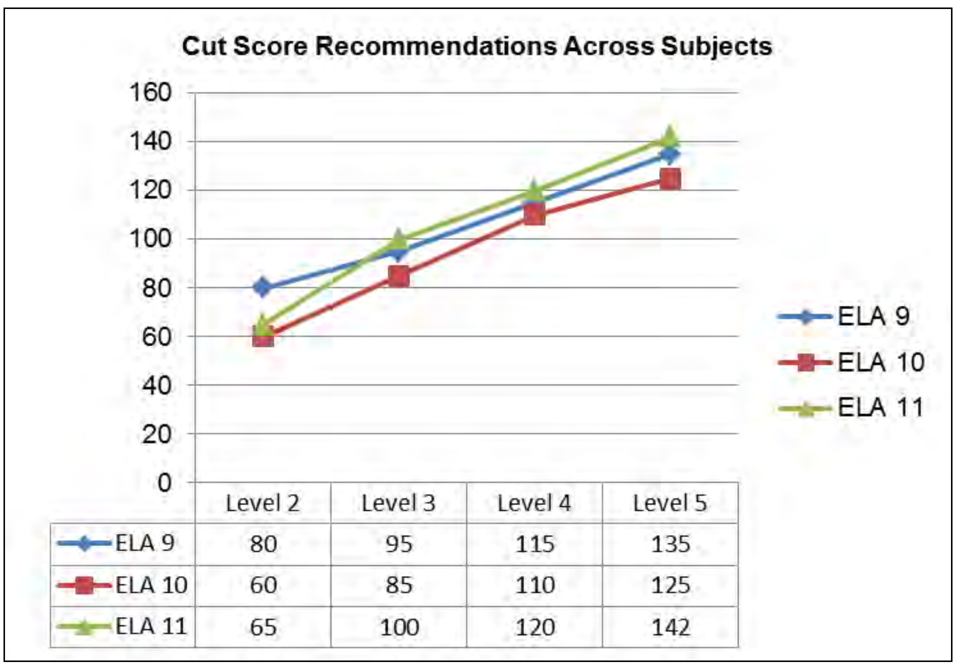
PARCC State Review of the Recommended High School Standards

After the standard-setting workshop is completed for the high school assessments, the results of the recommendations will be made available through webinars to PARCC states for their review. The webinars will need to take place during the first week of August (8/3/15-8/7/13) so that the states' feedback can be provided to the PARCC governing board and the ACCR for sign-off on the high school standards by 8/14. Three separate webinars will be provided by Pearson during this week to address the three separate high school course pathways: one for the ELA assessments, one for the traditional mathematics assessments, and one for the integrated mathematics assessments.

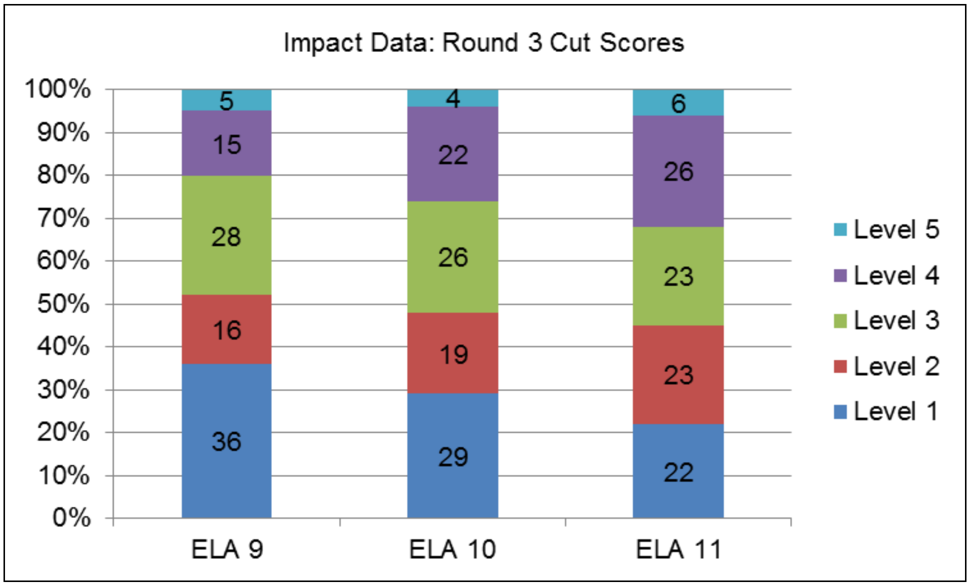
During the webinar, information will be provided to participants about the standard setting process, the recommended cut scores, and empirical results including impact data, vertical scale data if available, and information about how the cut scores line up with the external validity study results. Much of the data presented will be the same as what was presented during the vertical articulation segment of the standard-setting workshop as discussed in Section V.E.6. The figure below shows one way of presenting the Round 3 cut score recommendations across the levels. This would be provided for each high school assessment.

Round 3 Cut Score Recommendations				
Performance Standard	**Level 2**	**Level 3**	**Level 4**	**Level 5**
Minimum Page Number				
Maximum Page Number				
Mean Page Number				
Median Page Number				

In order to review the recommended cut scores across assessments within a course pathway, the following figure shows one possible representation to aid in the discussion of the reasonableness of the cut scores across the assessments.



The impact data associated with the Round 3 recommended cut scores will also be shared. The figure below shows one way to represent the impact data across grade levels within a subject.



Additionally, it will be important to show the Round 3 cut score information alongside the external validity study results to aid in the evaluation of the reasonableness of the recommendations in light of external validity evidence. One possible way to present this information in a user-friendly format is shown in the following figure.

PARCC Raw Score	PARCC Theta	Impact Data	Benchmark Studies Impact Data	Judgment Studies Cut Scores	Round 3 Cut Scores Level 2	Round 3 Cut Scores Level 3	Round 3 Cut Scores Level 4	Round 3 Cut Scores Level 5
0	-4.091	100						
⋮	⋮	⋮						
70	0.478	80			Minimum			
71	0.514	79			Median/Mean			
72	0.523	75						
73	0.542	71			Maximum			
74	0.633	66						
75	0.794	65				Minimum		
76	0.898	62						
77	0.898	58				Median		
78	1.024	55				Mean		
79	1.128	54						
80	1.204	53						
81	1.454	51		2YT Composition		Maximum		
82	1.756	50					Minimum	
83	1.808	47		2YT Technical				
84	2.068	45		2YT Literature				Minimum
85	2.309	41					Median	
⋮	⋮	⋮		4YO Composition			Mean	
110	2.432	31						

PARCC Raw Score	PARCC Theta	Impact Data	Benchmark Studies Impact Data	Judgment Studies Cut Scores	Round 3 Cut Scores Level 2	Round 3 Cut Scores Level 3	Round 3 Cut Scores Level 4	Round 3 Cut Scores Level 5
111	2.812	30	NAEP Proficient (30)	4YS Composition			Maximum	Median
0	-4.091	100						
⋮	⋮	⋮						
70	0.478	80			Minimum			
71	0.514	79			Median/Mean			
72	0.523	75						
73	0.542	71			Maximum			
74	0.633	66						
75	0.794	65				Minimum		
76	0.898	62						
77	0.898	58				Median		
78	1.024	55				Mean		
79	1.128	54						
80	1.204	53						
81	1.454	51		2YT Composition		Maximum		
82	1.756	50					Minimum	
83	1.808	47		2YT Technical				
84	2.068	45		2YT Literature				Minimum
85	2.309	41					Median	
⋮	⋮	⋮		4YO Composition			Mean	

PARCC Raw Score	PARCC Theta	Impact Data	Benchmark Studies Impact Data	Judgment Studies Cut Scores	Round 3 Cut Scores Level 2	Round 3 Cut Scores Level 3	Round 3 Cut Scores Level 4	Round 3 Cut Scores Level 5
110	2.432	31						
111	2.812	30	NAEP Proficient (30)	4YS Composition			Maximum	Median
112	2.820	27						
113	2.836	24						Mean
114	2.952	20						
115	2.973	18						
116	3.023	17						
117	3.159	15		4YO & 4YS Literature				
118	3.404	14						
⋮	⋮	⋮						
130	3.719	10						
131	3.728	7						Maximum
132	3.733	5						
133	4.633	3						
⋮	4.678	2						
150	4.997	1						

At the end of each high school review webinar, a question and answer session can be conducted to gather qualitative data depending on PARCC's preference. Finally, Pearson will work with PARCC to devise an instrument to collect feedback from states on the recommended standards. For example, an online survey could be provided to panelists.

PARCC Governing Board and ACCR Review of the Recommended High School Standards

Following the state review of the recommended high school PARCC performance standards, a second post-policy review will be conducted with the PARCC governing board and the ACCR. During this meeting, recommendations from the standard-setting workshop and information from the state post-policy review will be provided. In this meeting, the high school standards will be reviewed to verify that the standards are reasonable in light of empirical validity evidence and policy concerns across grades and subjects and meet the key purpose of the PARCC high school assessments—namely to determine college and career readiness for high school students.

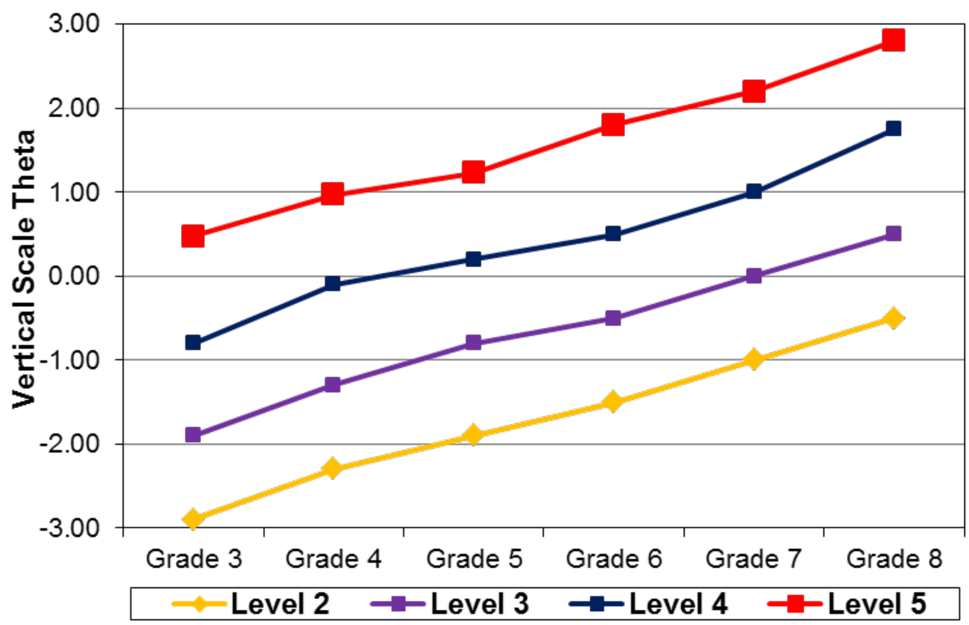
During this post-policy panel, the same figures and tables that will be used in the state post-policy review will also be provided in this post-policy review, along with additional information from the state review process. Additionally, Pearson will create an Excel workbook which will recompute and display updated impact data and vertical scale diagrams to allow on-the-fly adjustments to cut scores during this meeting as they are discussed. This will allow the PARCC governing board and ACCR to see the impact of possible cut score changes on student performance. This on-the-fly workbook method has been used extensively by Pearson in support of many states' standard setting processes.

PARCC State Review of the Recommended Grade 3-8 Standards

As with the high school assessments, there are two weeks between the grade 3-8 standard setting workshops and the date the standards are scheduled for approval by the PARCC governing board. Because of this short timeline, a state review of the recommended PARCC performance standards should occur during the week following the 3-8 standard-setting workshops. Standard-setting results for the 3-8 assessments will be made available through a webinar to PARCC states for their review. Four separate webinars will be provided by Pearson during this week: one for the grades 7 and 8 ELA assessments, one for the grades 3-6 ELA assessments, one for the grades 7 and 8 mathematics assessments, and one for the grades 3-6 mathematics assessments. During the webinar, information will be provided to participants about the standard setting process, the recommended cut scores, and empirical results including impact data, vertical scale data if available, and information about how the cut scores line up with the external validity study results.

In addition to the information provided which mirrors what will be presented during the high school reviews (as discussed previously), the figure provides an illustrative example of how vertical alignment information could be provided during the post-policy state review for grades

3-8. In this example, the four standards are compared in the theta metric across grades 3-8. If PARCC desires, and in support of fully articulated standards, high school results can also be provided during the grades 7 and 8 webinars to show participants how the middle school standards line up with the approved high school standards.



A question and answer session can be conducted to collect qualitative information depending on PARCC's preference. Finally, the instrument that was developed to collect feedback for the high school state post-policy reviews will be used to collect feedback for the 3-8 post-policy reviews.

PARCC Governing Board Review of the Recommended Grade 3-8 Standards

Upon completion of the 3-8 webinars, a final post-policy review will be conducted with the PARCC governing board. The process will be similar to that used for the high school post-policy review including an evaluation of standards across grades and subjects. The approved high school standards can be used as information to evaluate the alignment between the high school assessments and the 3-8 assessments. Pearson will provide on-the-fly results for possible adjustments to the standards and implement changes as requested.

Schedule of Major Milestones for Review of Cut Scores

The figure below shows a tentative schedule of the major milestones associated with the review of cut scores. Highlighted rows (3, 8, and 11) represent the requested dates for the standard setting events. Bolded text (rows 7 and 14) represents the requested dates for the PARCC governing board and ACCR review.

	Major Milestones for Review of Cut Scores	Start Date	End Date
1	Tools and materials for adjusting High School cut scores on the fly to PARCC governing board and ACCR	7/3/2015	7/3/2015
2	Tools and materials for adjusting Grades 3-8 cut scores on the fly to PARCC governing board and ACCR	7/20/2015	7/20/2015
3	Standard Setting for High School Assessments	7/27/2015	7/31/2015
4	PARCC state review of the recommended ELA standards Grades 9, 10, 11	8/6/2015	8/6/2015
5	PARCC state review of the recommended Math standards Algebra I, Geometry, Algebra II	8/6/2015	8/6/2015
6	PARCC state review of the recommended Math standards Integrated I, Integrated II, Integrated III	8/7/2015	8/7/2015
7	**PARCC governing board and ACCR review of the recommended high school standards**	**8/1/2015**	**8/14/2015**
8	Standard Setting for Grades 7-8	8/17/2015	8/21/2015
9	PARCC state review of the recommended ELA standards Grades 7, 8*	8/23/2015	8/23/2015
10	PARCC state review of the recommended Math standards Grades 7, 8*	8/23/2015	8/23/2015
11	Standard Setting for Grades 3-6	8/24/2015	8/28/2015
12	PARCC state review of the recommended ELA standards Grades 3, 4, 5, 6	9/7/2015	9/7/2015
13	PARCC state review of the recommended Math standards Grades 3, 4, 5, 6	9/8/2015	9/8/2015
14	**PARCC governing board review of the recommended 3-8 standards**	**8/31/2015**	**9/11/2015**

V.E.8. Technical Documentation

Requirement

Response Requirements for Section V.E.8.

a) Descriptions of the approach and procedures to complete all the responsibilities/tasks specified in Section V.C.8

Deliverables for Section V.E.8.

a) Draft and final Standard Setting Technical Report
   i. The Contractor shall complete the draft manual within 3 months of approval of final cut scores. PARCC will provide feedback on the draft and the Contractor shall complete the final report within 6 weeks of the PARCC feedback or a date mutually-agreed upon by the Partnership and the Contractor

Response

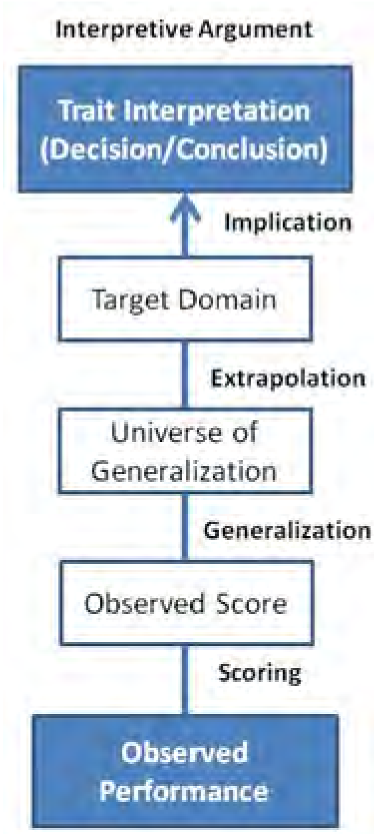
Technical Documentation of the Standard Setting Process and Results

Because of Pearson's experience integrating empirical data into standard setting and implementing Evidence Based Standard Setting (EBSS), Pearson understands how this standard-setting methodology provides evidence to the validity framework and has demonstrated this knowledge with Standard Setting Technical Reports for other assessments. The PARCC Standard Setting Technical Report will provide a comprehensive description of the standard-setting activities, from planning through results, and will explicitly document the ways in which the process contributes evidence to the validity framework of the assessments.

While the document will contain a great deal of information, to increase readability, it will include an Executive Summary, and it could be organized into the following major sections: Validity Framework, Standard Setting Design, Standard Setting Participants, Standard Setting Implementation, and Standard Setting Results.

Validity Framework

This section will provide a description and overview of the PARCC validity framework, specifically drawing from the relevant literature and the *Standards for Educational Measurement and Psychological Assessment* (AERA, APA, NCME, 1999). Also included will be an overview of Kane's argument-based approach to validity which outlines a process of moving from observed performance on assessments to decisions made from assessments as shown in the figure below (Kane, 2006). This section will describe each of the major steps described by Kane and summarize the way in which the PARCC standard setting lends evidence to the validity argument.



Also included in the validity framework section will be the uses, inferences, and assumptions of the PARCC assessment program, as these are central to the validity argument. Pearson will work closely with the PARCC governing board, operational working groups, the technical advisory committee and other relevant experts to document, edit, and review these inferences and assumptions. During the review of the performance level descriptors (PLDs) for standard setting, the PLDs will be considered specifically to enhance their alignment to the PARCC inferences and assumptions and to confirm they appropriately fit within the validity framework.

Each of the successive sections of the Standard Setting Technical Report will specifically address how that portion of the standard-setting process provides validity evidence and how that evidence fits within the validity argument. Within these descriptions will be differentiations between procedural and internal consistency evidence for validity.

Standard Setting Design

This section will outline the details of the EBSS design implemented for PARCC. Included in this section will be detailed descriptions of each of the empirical studies as well as information regarding how the studies were used within the standard-setting process such as informing the creation of reasonable ranges and PLD revisions. This section will focus on the advantages of using empirical data within the process and the way in which this contributes procedural evidence to the validity argument. For example, the use of empirical data to create reasonable ranges enhances the process by establishing boundaries in which to set performance standards. The reasonable ranges encourage reasonable cut scores which are an essential aspect of procedural validity evidence (Kane, 2001).

Standard Setting Participants

This section will describe the procedures followed to recruit participants as well as the composition of each of the standard-setting panels. Included will be descriptions regarding each of the characteristics that were considered in recruiting and creating the panels, such as the representativeness by demographics, member states, stakeholder group (educator, parent, community member), special populations, geographic region, and educator experience. The qualifications and representativeness of the standard-setting panels will contribute to the procedural evidence of the validity argument. Specifically, because standard setting is a judgment-based process, it is critical that the judgments come from well-qualified panelists. Well-formed and experienced panels enhance the procedural validity evidence of the standard-setting process (Kane, 2001).

Standard Setting Implementation

The implementation section will provide details of the operational standard-setting workshops. This section will describe, step-by-step, the proceedings of the workshops including the materials used, the process to review and validate the PLDs, the training on the assessments and the standard-setting method, the creation of the borderline student descriptors, the procedures for each round of ratings, and the feedback provided to the panels. Also included will be a description of the steps taken to create aligned performance standards, specifically in regards to vertical articulation and consistency across the high school mathematics assessments. This section will contain information regarding how the standard-setting implementation provides procedural validity evidence and the steps taken to enhance the internal consistency of the resulting cut scores. For example, documentation regarding the adherence to the planned and well-defined standard-setting method provides one source of procedural evidence (Kane, 2001).

Standard Setting Results

This section will provide the results of the standard setting workshops including the round-by-round cut score recommendations, the results of the vertical articulation, the feedback collected from the evaluation survey, feedback from the PARCC state review process, and any adjustments made to cut scores during the post-policy meeting for policy considerations. Included in this section will be both quantitative and qualitative summaries of the standard-setting data. The standard-setting results will be discussed in regards to procedural and internal consistency evidence for the validity argument. Specifically, standard error statistics, computed using the standard-setting data, provide internal consistency evidence (Kane, 2001). Decision consistency and accuracy statistics will also be included in this section.

Summary

Pearson will create a Standard Setting Technical Report that provides a thorough description of the PARCC standard setting process and specifically addresses how this process contributes substantial evidence to the PARCC validity framework. Pearson will provide this documentation in the requested timeframe allowing needed time for input from stakeholder groups and experts identified by PARCC.

References

American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). (1999). Standards for educational and psychological testing. Washington, DC: AERA.

Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed.) (pp.17-64). Westport, CT: Greenwood Publishing.

Kane, M. (2001). So much remains the same: Conception and status of validation in setting standards. In G.J. Cizek (Ed.). Setting performance standards: Theory and applications (pp.53-88) Mahwah, NJ: Lawrence Erlbaum.

V.E.9. Quality Control

Requirement

Response Requirements for Section V.E.9.

a) Description of the approach and procedures to complete all the responsibilities/tasks specified in Section V.E.9

b) Description of the steps that will be taken to protect the security of the test items used for standard setting and standard setting results

Deliverables for Section V.E.9.

a). Quality control specifications that describe in detail all of the steps to be implemented to demonstrate to the Partnership that data input and resulting reports are correct and free of security breaches

Response

Quality Control for Data Input and Resulting Reports

Pearson will implement quality control procedures to achieve accuracy in the input, processing, and reporting of data associated with the standard setting. In this section, we provide details of the steps that will be taken to establish these quality control procedures.

Specifications and Quality Processes for Data Analysis and Reporting

Pearson's proposed standard setting approach uses tools and associated data analyses systems that we have developed and refined over a number of standard settings. In particular, we have written SAS analysis programs that efficiently interface with the Qwizdom Response System to input and summarize judgments transmitted by panelists using these devices. In our experience, use of the Qwizdom devices results in more rapid and more accurate input of the standard setting judgment data.

As previously discussed, data analysis programs used to generate the feedback data and compute the recommended cut scores will be rigorously tested prior to the standard setting meeting. The quality control processes will include running test data through the program to verify output, and the use of checklists to document the application of reasonableness checks performed on program outputs.

Within the standard setting plan, Pearson will document the specifications and quality processes for data analysis and reporting. Within the plan, we will include descriptions of the programs to be used, the analyses to be conducted, conventions for naming and managing input and output data sets, descriptions of graphical and tabular outputs, and process checklists.

Pearson will implement tools and data analyses as part of the practice standard setting, which will provide an opportunity for additional quality control checks prior to the operational standard setting. These may include independent verification of program outputs and calculations.

Standard Setting Audit

Pearson will collaborate with and support the three technical advisory committee members and four PARCC state representatives identified to perform quality audits of the standard setting workshop. We will work with these individuals to identify their roles and responsibilities and how they will participate in the standard setting event. We will work with them to develop materials that they can use in observing and evaluating the procedures followed during the standard setting. These may include checklists and, if requested, raw data for independent analysis.

As part of the practice standard setting described above, Pearson will involve the technical advisory committee and PARCC state representatives to perform initial quality audits. The audit results for the practice standard setting will be used to guide any needed process refinements prior to the operational standard setting.

Security of Standard Setting Materials

Pearson will implement a number of steps to protect the security of the test items used for standard setting and standard setting results. In keeping with similar security procedures associated with item reviews, Pearson will adhere to the review meeting requirements in Section V.A. Test Development of the response for proposals.

Based on our experience with other online testing programs, security will be enhanced by the use of computers as a basis for viewing the test items. However, paper materials containing confidential data or information will be checked out at the beginning of each day, checked back in at the end of the day, and stored in a secure location during off hours.

Standard setting participants that have contact with test items, including those in the various studies to inform the standard setting process, will sign confidentiality agreements acknowledging that they will not share information about the test items outside of the meeting or research study in which they are participating.

Component 6: Program Management

V.F.1 Program Management Plan

Requirement

V.F.1.A. Program Management Approach

Response Requirements for Section V.F.1.A.

a) Offeror’s proposal shall include a response to the requirements specified in Section V.F.1.A.

Deliverables for Section V.F.1.A.

a) Deliver Annual Program Management Plan

Response

Our experience has shown us that correct application of project management skills, tools, and techniques enhance the likelihood of success for a wide range of projects—including the PARCC Operational Assessments. We will use these skills in the development of our Program Management Plan.

This annual highly detailed Program Management Plan will be the outline for delivery of the program including processes and methodologies for the following:

- Staff management
- Subcontractor Management
- Communication Management
- Scope/Change Management
- Quality Management
- Cost Management
- Risk Management
- Schedule Management

We recognize that transparency, timely communication, and problem solving are important to PARCC, and these characteristics are at the foundation of the Pearson program management team, with each member carefully selected for PARCC. Pearson will make the

Program Management Plan, our schedule, risk logs, action items, and decision point documents available on a mutually agreed upon portal, which may be a SharePoint website or other web-based system easy for PARCC and the Partnership Manager to use.

Requirement

V.F.1.B. Staff Management

1. Provisions Governing the Work of the Contractor
2. Staff Management Requirements

Response Requirements for Section V.F.1.B.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.1.B.

Response

As the prime contractor, Pearson will be responsible for providing a strong management structure to initiate tasks and to monitor on-time completion throughout the contract period. Below are executive leaders that will be available for internal consulting as well as for customer meetings, as needed. Several members of our executive team have attended and presented at PARCC Governing Board and Technical Advisory Meetings, including Doug Kubach, Dr. Jon Twing, Dr. Denny Way, and Pat Kramer. These executive leaders will support the key personnel proposed for this project.

PARCC Operational Assessments: Pearson Executive Management Team		
Name	**Title**	**Role and Responsibilities**
Doug Kubach	President, Assessment and Instruction	Executive Oversight Overall program
Walter Sherwood	President, State Services	Executive Oversight, Program delivery
Alistair Van Moere	President Knowledge Technologies	Executive Oversight, Automated scoring
Jon S. Twing	Executive Vice President and Chief Measurement Officer, Assessment and Instruction	Executive Oversight Psychometrics, research, content, and technology
Walter (Denny) Way	Senior Vice President, Measurement Services	Executive Oversight Psychometrics, research and content
Julie Miles	Vice President, Measurement Services	Executive Oversight Psychometrics and research and content
Kimberly O'Malley	Senior Vice President, Research & Innovation Network	Executive Oversight Research & Innovation Network
Wayne Ostler	Vice President, Digital Content and Measurement Systems	Executive Oversight Digital content delivery
KJ Singh	Senior Vice President and Chief Technology Officer	Executive Oversight, Technology

PARCC Operational Assessments: Pearson Executive Management Team		
Name	**Title**	**Role and Responsibilities**
Brendan Kealey	Vice President, Product Development	Executive Oversight, Product development
Randy Schuessler	Vice President, Assessment Solutions Development and Delivery	Executive Oversight, Assessment development and delivery
Stephanie Rompot	Director, Program Technology Management	Executive Oversight, Program technology
Jim Hummer	Senior Vice President and Chief Quality Officer	Executive Oversight, Quality
Sue Ann Averitte	Vice President, Quality Assurance and Continuous Improvement	Executive Oversight Quality assurance
Kate Minette	Senior Vice President, Operations and Scoring	Executive Oversight Performance Scoring
Anne Parmley	President, National Services	Executive Oversight Overall program and functional management of program management staff
Pat Kramer	Vice President, National Services	Executive Oversight for the overall program and functional management of program management staff

Given ETS's significant role on current PARCC programs and its proposed scope of work for PARCC Operational Assessments, executive-level management for ETS is also provided in the following figure. Management staffing for Caveon, Measured Progress, and WestEd is described later in this section.

PARCC Operational Assessments: ETS Executive Management Team		
Name	**Title**	**Role and Responsibilities**
Stephen Lazer	Vice President, Student and Teacher Assessments	Executive Oversight
John Oswald	Vice President and General Manager, K–12 Student Assessment Programs	Executive Oversight
John Mazzeo	Vice President, Statistical Analysis, Data Analysis, and Psychometric Research	Executive Oversight
Marisa Farnum	Vice President, Assessment Development	Executive Oversight
Patricia Klag	General Manager, Assessment Development	Executive Oversight
Fred McHale	General Manager, Assessment Development	Executive Oversight

To develop the PARCC Operational Assessments, we have assembled an experienced, high-performing team chosen from across Pearson and subcontractor organizations based on specific RFP requirements and relevant skill sets to best serve PARCC. Together, our team provides expertise in the areas of item and test development, online testing, packaging and distribution, scanning, automated and professional scoring, research, psychometrics, report usability and design, meeting planning and program management.

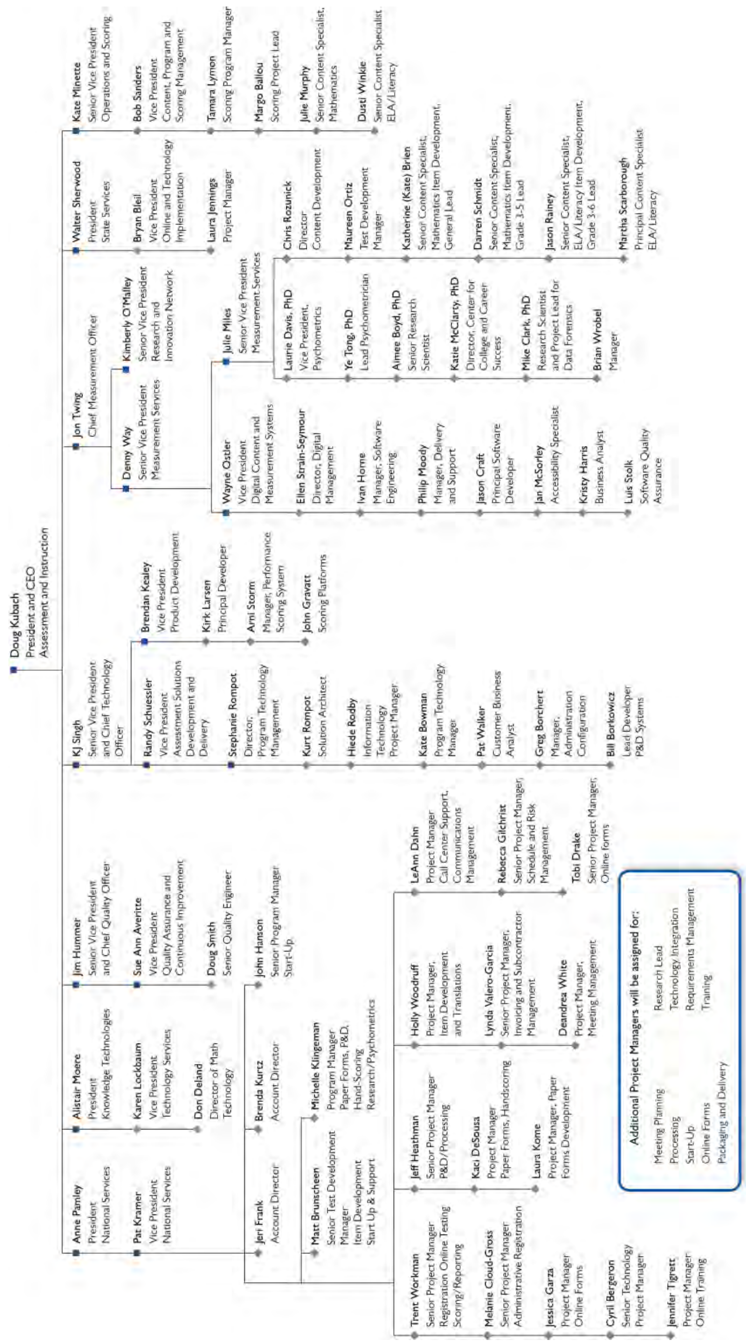
The collaboration's structure is based on the following division of labor across organizations:

Expert Team for PARCC Assessments	
Component	**Primary Responsibility**
Test Development	Pearson, ETS, WestEd
Assessment Administration	Pearson
Psychometric Services	Pearson, ETS Measured Progress (Quality Control) Caveon (Data Forensics: Internet Monitoring)
Reporting	Pearson
Standard Setting	Pearson ETS, WestEd (Content Facilitators)
Program Management	Pearson

Lines of Authority

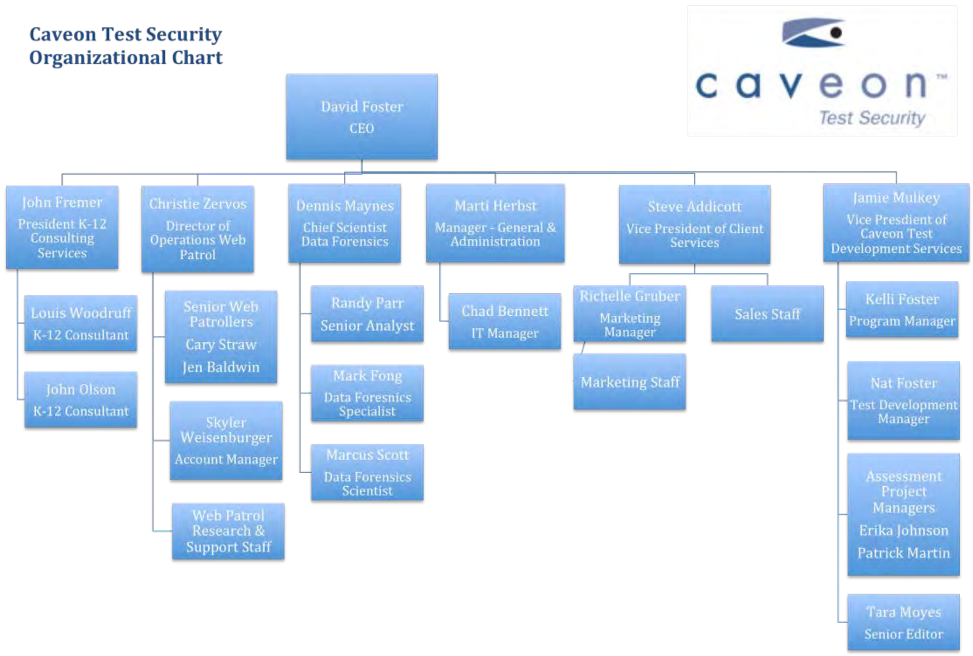
Following are organizational charts for Pearson, Caveon, ETS, Measured Progress, and WestEd.

Pearson



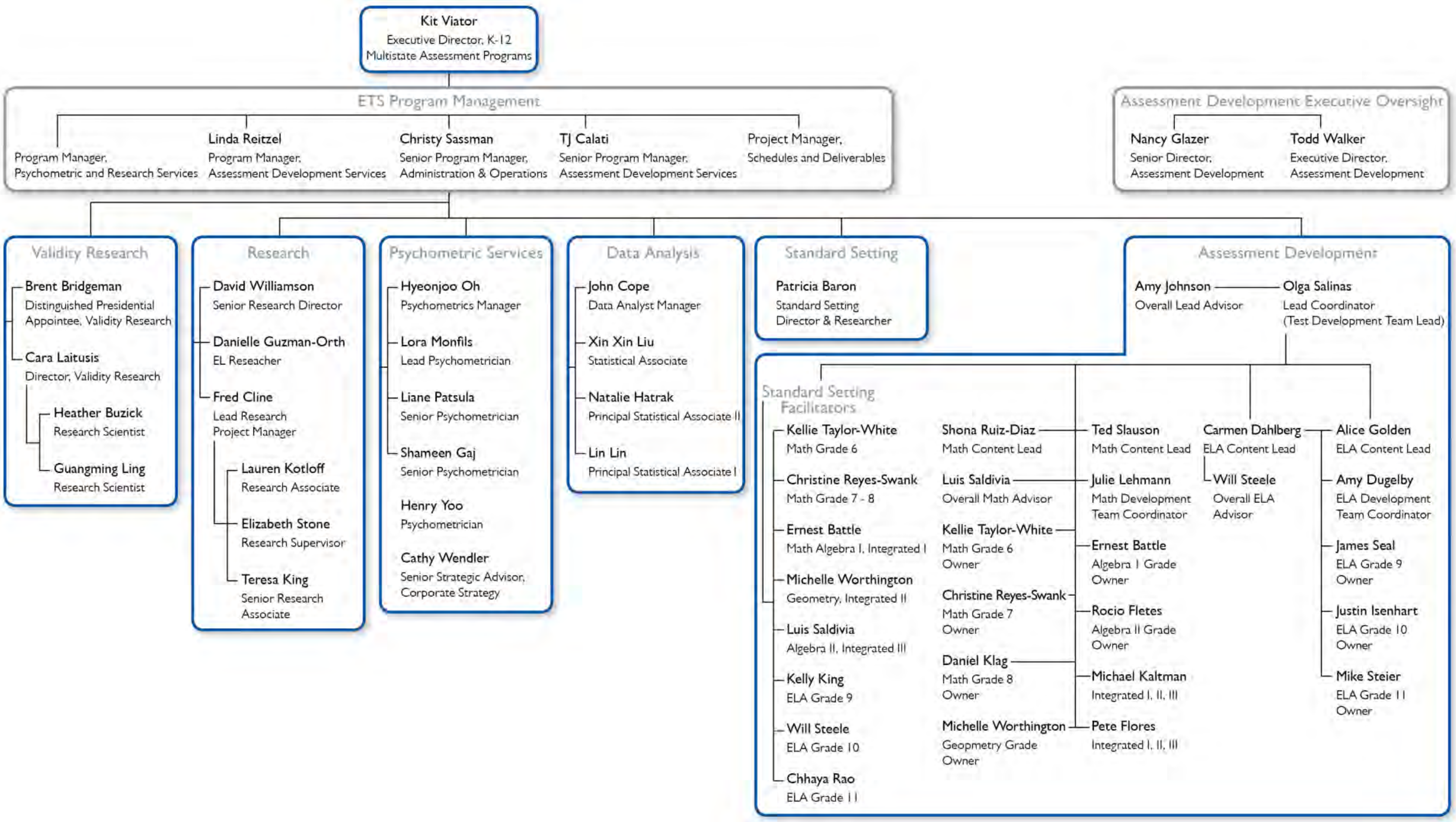
Doug Kubach
President and CEO
Assessment and Instruction
Anne Pamley
President
National Services
Pat Kramer
Vice President
National Services
Jeri Frank
Account Director
Matt Brunscheen
Senior Test Development Manager
Item Development
Start Up & Support
Alistair Moore
President
Knowledge Technologies
Karen Lochbaum
Vice President
Technology Services
Don Deland
Director of Math Technology
Brenda Kurtz
Account Director
Michelle Klingeman
Program Manager
Paper Forms, P&D,
Hand-Scoring
Research/Psychometrics
Jim Hummer
Senior Vice President
and Chief Quality Officer
Sue Ann Averitte
Vice President
Quality Assurance and
Continuous Improvement
Doug Smith
Senior Quality Engineer
John Hanson
Senior Program Manager
Start-Up
Trent Workman
Senior Project Manager
Registration Online Testing
Scoring/Reporting
Melanie Cloud-Gross
Senior Project Manager
Administrative Registration
Jessica Garza
Project Manager
Online Forms
Cyril Bergeron
Senior Technology
Project Manager
Jennifer Tigrett
Project Manager
Online Training
Jeff Heathman
Senior Project Manager
P&D/Processing
Kaci DeSousa
Project Manager
Paper Forms, Handscoring
Laura Kome
Project Manager, Paper
Forms Development
Holly Woodruff
Project Manager,
Item Development
and Translations
Lynda Valero-Garcia
Senior Project Manager,
Invoicing and Subcontractor
Management
Deandrea White
Project Manager,
Meeting Management
LeAnn Dahn
Project Manager
Call Center Support,
Communications
Management
Rebecca Gilchrist
Senior Project Manager,
Schedule and Risk
Management
Tobi Drake
Senior Project Manager,
Online forms
Additional Project Managers will be assigned for:
Meeting Planning
Processing
Start-Up
Online Forms
Packaging and Delivery
Research Lead
Technology Integration
Requirements Management
Training
KJ Singh
Senior Vice President
and Chief Technology
Officer
Randy Schuessler
Vice President
Assessment Solutions
Development and
Delivery
Stephanie Rompot
Director,
Program Technology
Management
Kurt Rompot
Solution Architect
Hiede Rodby
Information
Technology
Project Manager
Kate Bowman
Program Technology
Manager
Pat Walker
Customer Business
Analyst
Greg Borchert
Manager,
Administration
Configuration
Bill Borkowicz
Lead Developer
P&D Systems
Brendan Kealey
Vice President
Product Development
Kirk Larsen
Principal Developer
Arni Storm
Manager, Performance
Scoring System
John Gravatt
Scoring Platforms
Jon Twing
Chief Measurement Officer
Denny Way
Senior Vice President
Measurement Services
Kimberly O'Malley
Senior Vice President
Research and
Innovation Network
Wayne Ostler
Vice President
Digital Content and
Measurement Systems
Ellen Strain-Seymour
Director, Digital
Management
Ivan Horne
Manager, Software
Engineering
Philip Moody
Manager, Delivery
and Support
Jason Craft
Principal Software
Developer
Jan McSorley
Accessibility Specialist
Kristy Harris
Business Analyst
Luis Stolk
Software Quality
Assurance
Julie Miles
Senior Vice President
Measurement Services
Laurie Davis, PhD
Vice President,
Psychometrics
Ye Tong, PhD
Lead Psychometrician
Aimee Boyd, PhD
Senior Research
Scientist
Katie McClarty, PhD
Director, Center for
College and Career
Success
Mike Clark, PhD
Research Scientist
and Project Lead for
Data Forensics
Brian Wrobel
Manager
Walter Sherwood
President
State Services
Bryan Bleil
Vice President
Online and Technology
Implementation
Laura Jennings
Project Manager
Chris Rozunick
Director
Content Development
Maureen Ortiz
Test Development
Manager
Katherine (Kate) Brien
Senior Content Specialist,
Mathematics Item Development,
General Lead
Darren Schmidt
Senior Content Specialist,
Mathematics Item Development,
Grade 3-5 Lead
Jason Rainey
Senior Content Specialist,
ELA/Literacy Item Development,
Grade 3-6 Lead
Martha Scarborough
Principal Content Specialist
ELA/Literacy
Kate Minette
Senior Vice President
Operations and Scoring
Bob Sanders
Vice President
Content, Program and
Scoring Management
Tamara Lymon
Scoring Program Manager
Margo Ballou
Scoring Project Lead
Julie Murphy
Senior Content Specialist,
Mathematics
Dusti Winkle
Senior Content Specialist
ELA/Literacy

Caveon



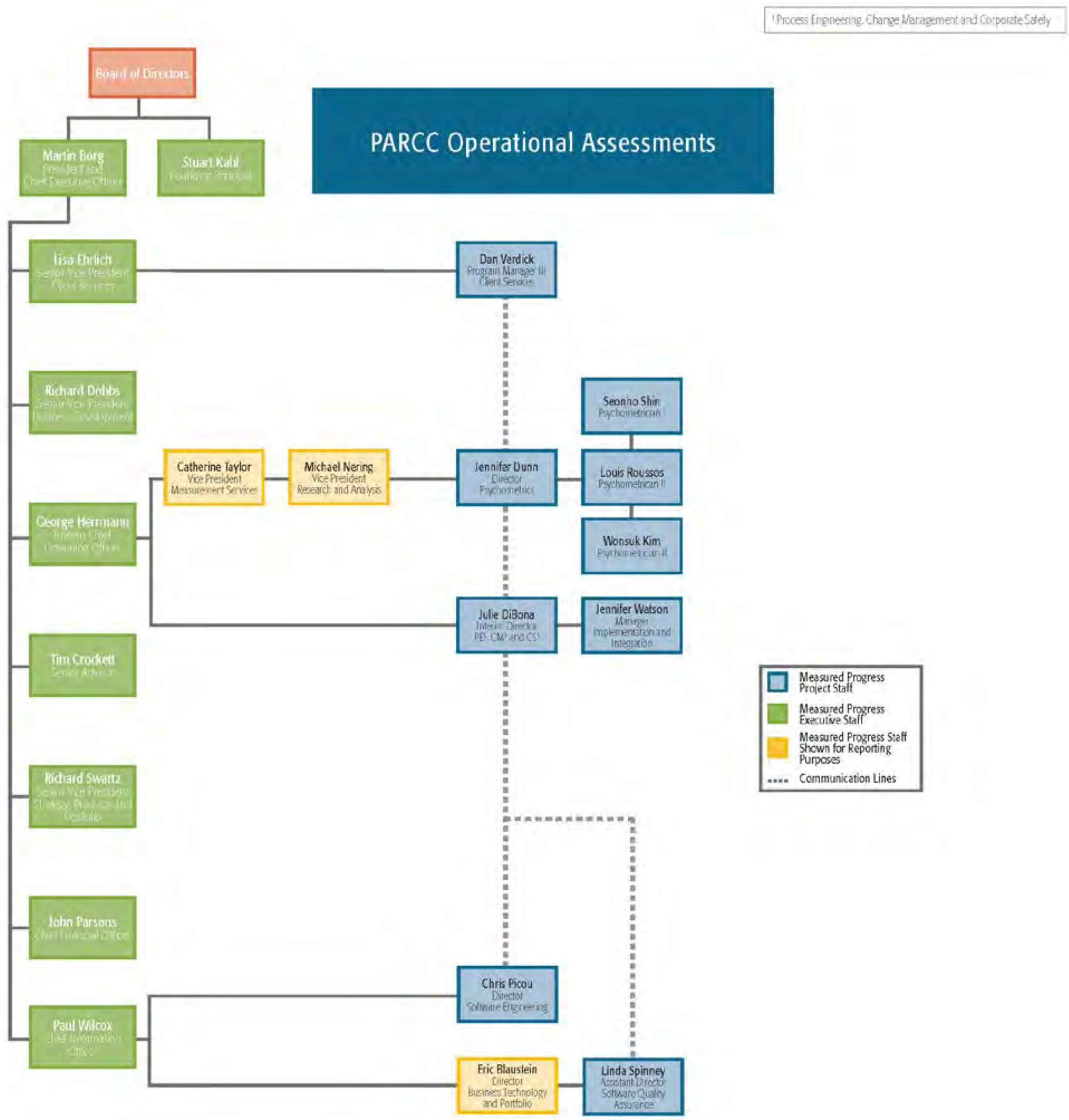
Caveon Test Security
Organizational Chart
caveon
Test Security
David Foster
CEO
John Fremer
President K-12 Consulting Services
Christie Zervos
Director of Operations Web Patrol
Dennis Maynes
Chief Scientist Data Forensics
Marti Herbst
Manager - General & Administration
Steve Addicott
Vice President of Client Services
Jamie Mulkey
Vice Presdient of Caveon Test Development Services
Louis Woodruff
K-12 Consultant
John Olson
K-12 Consultant
Senior Web Patrollers
Cary Straw
Jen Baldwin
Skyler Weisenburger
Account Manager
Web Patrol Research & Support Staff
Randy Parr
Senior Analyst
Mark Fong
Data Foresnics Specialist
Marcus Scott
Data Forensics Scientist
Chad Bennett
IT Manager
Richelle Gruber
Marketing Manager
Sales Staff
Marketing Staff
Kelli Foster
Program Manager
Nat Foster
Test Development Manager
Assessment Project Managers
Erika Johnson
Patrick Martin
Tara Moyes
Senior Editor

ETS

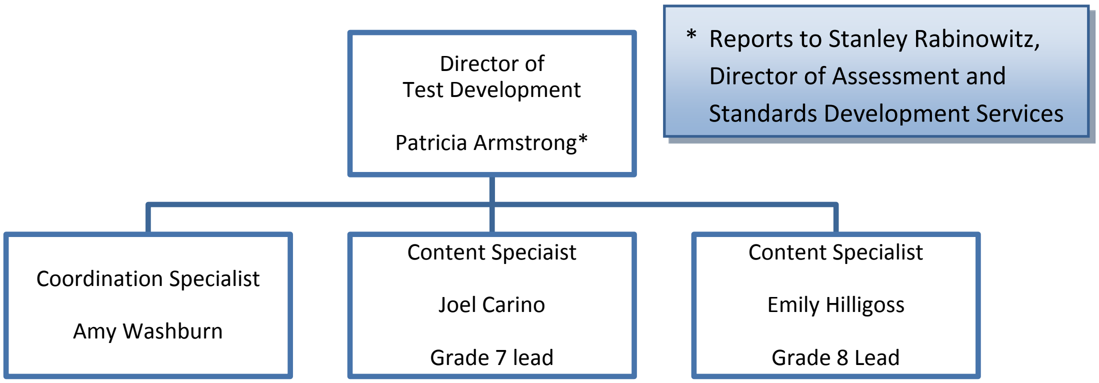

Kit Viator
Executive Director, K-12
Multistate Assessment Programs
ETS Program Management
Program Manager,
Psychometric and Research Services
Linda Reitzel
Program Manager,
Assessment Development Services
Christy Sassman
Senior Program Manager,
Administration & Operations
TJ Calati
Senior Program Manager,
Assessment Development Services
Project Manager,
Schedules and Deliverables
Assessment Development Executive Oversight
Nancy Glazer
Senior Director,
Assessment Development
Todd Walker
Executive Director,
Assessment Development
Validity Research
Brent Bridgeman
Distinguished Presidential
Appointee, Validity Research
Cara Laitusis
Director, Validity Research
Heather Buzick
Research Scientist
Guangming Ling
Research Scientist
Research
David Williamson
Senior Research Director
Danielle Guzman-Orth
EL Reseacher
Fred Cline
Lead Research
Project Manager
Lauren Kotloff
Research Associate
Elizabeth Stone
Research Supervisor
Teresa King
Senior Research
Associate
Psychometric Services
Hyeonjoo Oh
Psychometrics Manager
Lora Monfils
Lead Psychometrician
Liane Patsula
Senior Psychometrician
Shameen Gaj
Senior Psychometrician
Henry Yoo
Psychometrician
Cathy Wendler
Senior Strategic Advisor,
Corporate Strategy
Data Analysis
John Cope
Data Analyst Manager
Xin Xin Liu
Statistical Associate
Natalie Hatrak
Principal Statistical Associate II
Lin Lin
Principal Statistical Associate I
Standard Setting
Patricia Baron
Standard Setting
Director & Researcher
Assessment Development
Amy Johnson
Overall Lead Advisor
Olga Salinas
Lead Coordinator
(Test Development Team Lead)
Standard Setting
Facilitators
Kellie Taylor-White
Math Grade 6
Christine Reyes-Swank
Math Grade 7 - 8
Ernest Battle
Math Algebra I, Integrated I
Michelle Worthington
Geometry, Integrated II
Luis Saldivia
Algebra II, Integrated III
Kelly King
ELA Grade 9
Will Steele
ELA Grade 10
Chhaya Rao
ELA Grade 11
Shona Ruiz-Diaz
Math Content Lead
Luis Saldivia
Overall Math Advisor
Kellie Taylor-White
Math Grade 6
Owner
Christine Reyes-Swank
Math Grade 7
Owner
Daniel Klag
Math Grade 8
Owner
Michelle Worthington
Geopmetry Grade
Owner
Ted Slauson
Math Content Lead
Julie Lehmann
Math Development
Team Coordinator
Ernest Battle
Algebra I Grade
Owner
Rocio Fletes
Algebra II Grade
Owner
Michael Kaltman
Integrated I, II, III
Pete Flores
Integrated I, II, III
Carmen Dahlberg
ELA Content Lead
Will Steele
Overall ELA
Advisor
Alice Golden
ELA Content Lead
Amy Dugelby
ELA Development
Team Coordinator
James Seal
ELA Grade 9
Owner
Justin Isenhart
ELA Grade 10
Owner
Mike Steier
ELA Grade 11
Owner

PEARSON

Measured Progress



WestEd



Key Personnel Staffing

The following figures highlight key personnel from Pearson, Caveon, ETS, Measured Progress, and WestEd, proposed for the PARCC Operational Assessments program team. For each person we have included their name, title/project role, and percentage of time allocation. We have provided costs for new staff to be hired in addition to the key personnel included below. Unless otherwise noted, the anticipated timeframe of their involvement is indicated in terms of an average full time equivalent percentage across years. For some ETS staff, time allocations are shown separately for year one and the average across years two through four.

Pearson Key Personnel		
Name	**Title/Role**	**Time**
Program Management		
Jeri Frank	Account Director	100%
Brenda Kurtz	Account Director	100%
Matt Brunscheen	Program Manager, Item Banking and Item Development	100%
Michelle Klingeman	Senior Program Manager, Operational Startup	100%
Trent Workman	Program Manager, PearsonAccess, TestNav, Scoring and Reporting	100%
Monica Lyons	Program Manager, Customer Service, Paper Forms, Packaging and Distribution, Handscoring, Psychometrics, Standard Setting, and Research	100%
Cyril Bergeron	Senior Technical Project Manager, Scoring	100%
Melanie Cloud-Gross	Senior Project Manager, PearsonAccess	100%
LeAnn Dahn	Project Manager, Call Center and Communications Management	100%
Kaci DeSousa	Project Manager, Paper Forms and Handscoring	100%

Pearson Key Personnel		
Name	**Title/Role**	**Time**
Tobi Drake	Senior Project Manager, Online Forms	100%
Jessica Garza	Project Manager, Online Forms	100%
Rebecca Gilchrist	Senior Project Manager, Schedule and Risk Management	100%
Jeff Heathman	Senior Project Manager, Packaging, Distribution, and Processing	100%
Laura Kome	Project Manager, Paper Forms Development	100%
Jennifer Tigrett	Project Manager, Online Training	100%
Lynda Valero-Garcia	Senior Project Manager, Invoicing and Subcontractor Management	100%
Deandrea White	Project Manager, Meeting Management	100%
Holly Woodruff	Project Manager, Item Development and Translations	100%
Content Development		
Adrian Rivera	Test Development Manager	85%
Martha Scarborough	Principal ELA Content Specialist, ELA/Literacy Item Development, General Lead	100%
Jason Rainey	Senior Content Specialist, ELA/Literacy Item Development, Grade 3–6 Lead	100%
Katherine (Kate) Brien	Senior Content Specialist, Mathematics Item Development, General Lead	100%
Darren Schmidt	Senior Content Specialist, Mathematics Item Development, Grade 3–5 Lead	100%
Knowledge Technologies		
Karen Lochbaum	Vice President, Technology Services	10%
Don Deland	Director of Math Technology	10%
Software Technology Services		
Ellen Strain-Seymour	Director, Digital Management	15%
Ivan Horne	Manager, Software Engineering	15%
Philip Moody	Manager, Delivery and Support	20%
Jason Craft	Principal Software Developer	10%
Jan McSorley	Accessibility Specialist	15%
Kristy Harris	Business Analyst	20%
Luis Stolk	Software Quality Assurance	20%
Kate Bowman	Program Technology Manager	100%
Hiede Rodby	Information Technology Project Manager	100%
Pat Walker	Customer Business Analyst	100%
Kirk Larson	Principal Developer	75%
Kurt Rompot	Solution Architect	50%
Greg Borchert	Manager, Administration Configuration	50%
Arni Storm	Manager, Performance Scoring System	75%
John Gravatt	Manager, Technology Strategy	50%
Bill Borkowicz	Lead Developer, Packaging and Distribution System	50%

Pearson Key Personnel		
Name	**Title/Role**	**Time**
Bryan Bleil	Vice President, Online & Technology Implementation	20%
Laura Jennings	Project Manager	100%
Performance Scoring		
Bob Sanders	Vice President, Content, Program, and Scoring Management	25%
Tamara Lyman	Scoring Program Manager	100%
Margo Ballou	Scoring Project Lead	100%
Julie Murphy	Senior Content Specialist, Mathematics	100%
Dusti Winkie	Senior Content Specialist, ELA/L	100%
Psychometric Services and Research		
Laurie Davis	Vice President, Psychometrics	10%
Ye Tong	Lead Psychometrician	10%
Katie McClarty	Director, Center for College and Career Success	15%
Aimee Boyd	Senior Research Scientist	50%
Mike Clark	Research Scientist and Project Lead for Data Forensics	50%
Brian Wrobel	Manager	50%
Quality Assurance and Continuous Improvement		
Douglas Smith	Senior Quality Engineer	80%

Caveon Key Personnel		
Name	**Title/Role**	**Time**
Web Monitoring		
Christie Zervos	Director of Operations, Web Patrol Service	50%

ETS Key Personnel		
Name	**Title/Role**	**Time**
Program Management		
Kit Viator	Executive Director	15%
Sandra Wheatley	Program Manager	50%
Christy Sassman	Program Manager	35/50%
TJ Calati	Project Manager	75%
Program Manager	Psychometric and Research Services	50%
Jan Koekemoer	Program Manager, Schedules and Deliverables	50%
Validity Research		
Brent Bridgeman	Distinguished Presidential Appointee, Validity Research	>1%
Cara Laitusis	Director, Validity Research	5%
Heather Buzick	Research Scientist	>1%

ETS Key Personnel		
Name	**Title/Role**	**Time**
Guangming Ling	Research Scientist	5%
Research		
David Williamson	Senior Research Director	5%
Danielle Guzman-Orth	ELA/L Researcher	2%
Fred Cline	Lead Research Project Manager	15%
Lauren Kotloff	Research Associate	10%
Elizabeth Stone	Research Supervisor	20%
Teresa King	Senior Research Associate	10%
Psychometric Services		
Hyeonjoo Oh	Psychometrics Manager	5%
Lora Monfils	Lead Psychometrician	30/100%
Terran Brown	Senior Psychometrician	30/100%
Shameem Gaj	Senior Psychometrician	30/50%
Henry Yoo	Psychometrician	30/50%
Cathy Wendler	Strategic Advisor	29/17%
Data Analysis		
John Cope	Data Analyst Manager	15/10%
Xin Xin Liu	Statistical Associate	25/100%
Lin Lin	Principal Statistical Associate I	25/100%
Natalie Hatrak	Principal Statistical Associate II	20/60%
Standard Setting		
Patricia Baron	Standard Setting Director & Researcher	2%
Kelly Van Houten-King	English Language Arts Grade 9 Standard Setting Facilitator	10%
Chhaya Rao	English Language Arts Grade 11 Standard Setting Facilitator	10%
Assessment Development		
Nancy Glazer	Senior Director, Assessment Development	50%
Todd Walker	Executive Director, Assessment Development	10%
Amy Johnson	Overall Lead Advisor	20/50%
Olga Salinas	Lead Coordinator	35/50%
Carmen Dahlberg	ELA Content Lead	95%
Alice Golden	ELA Content Lead	40/95%
Will Steele	Overall ELA Advisor/ELA Development Team Coordinator	50/25%
James Seal	ELA Grade 9 Owner	65/95%
Justin Isenhart	ELA Grade 10 Owner	65/95%
Mike Steier	ELA Grade 11 Owner	65/95%
Shona Ruiz-Diaz	Math Content Lead	60/95%
Ted Slauson	Math Content Lead	75/95%
Luis Saldivia	Overall Math Advisor	25%

ETS Key Personnel		
Name	**Title/Role**	**Time**
Julie Lehmann	Math Development Team Coordinator	50%
Kellie Taylor-White	Math Grade 6 Owner	65/95%
Christine Reyes-Swank	Math Grade 7 Owner	65/95%
Daniel Klag	Math Grade 8 Owner	65/95%
Michelle Worthington	Geometry Grade Owner	65/95%
Ernest Battle	Algebra I Grade Owner	65/95%
Rocio Fletes	Algebra III Grade Owner	65/95%
Michael Kaltman	Integrated I, II, III	65/95%
Pete Flores	Integrated I, II, III	65/95%

Measured Progress Key Personnel		
Name	**Title/Role**	**Time**
Program Management		
Dan Verdick	Program Manager III	15%
Psychometrics		
Jennifer Dunn	Director	10%
Louis Roussos	Psychometrician II	33%
Wonsuk Kim	Psychometrician II	33%
Seonho Shin	Psychometrician I	33%

WestEd Key Personnel		
Name	**Title/Role**	**Time**
Test Development		
Patricia Armstrong	Director of Test Development	20%
Amy Washburn	Coordination Specialist	20%
Joel Carino	Content Specialist, Content Lead—Grade 7	40%
Emily Hilligoss	Content Specialist, Content Lead—Grade 8	40%

Operational Assessment Team. Each organization has chosen personnel with the experience and skills to successfully develop the PARCC Operational Assessment.

If a change in key personnel becomes a possibility, Pearson and our subcontractors, will provide the rationale for the change, the replacement personnel's resume, and a detailed transition plan when seeking written approval from PARCC for the change in key personnel.

Other Supporting Materials: Key Personnel Vitae includes one-page vitae for each key personnel named by Pearson and our subcontractors. Following are cameos that indicate relevant educational background and/or professional experience in their area of expertise.

Pearson Key Personnel

Pearson Program Management

To provide continuity and executive oversight to PARCC programs, **Pat Kramer** will provide management support to the PARCC Operational Assessments team. While not named as key personnel with specific percentage of time allocated, Ms. Kramer is available as needed for internal and external discussion and to escalate issues. She plays key roles in our PARCC Assessment Development and Assessment Administration contracts, and understands key dependencies, handoffs, and priorities for PARCC. Instead of building out discrete teams for each individual contract we are augmenting key staff to the existing PARCC Program Team so that the information and communication flows across contracts and is lock-step with PARCC's expectations, schedule, and changes as they occur.

Jeri Frank, Account Director. Jeri Frank (and Brenda Kurtz) will be PARCC's primary point of contact around contractual activities and will oversee the project. Ms. Frank and Ms. Kurtz will coordinate communications between PARCC and our PARCC Operational Assessment team, which includes subcontractor staff. Due to the aggressive schedule, it will be important for PARCC to have clean lines of communication. Ms. Frank has more than 15 years of program management experience, including 10 years with large-scale, high stakes assessments, and nearly two years of experience on PARCC.

In previous roles, she was responsible for fulfilling Pearson's contract providing test administration services in support of the ADP Algebra I End-of-Course Exam, the Florida Comprehensive Assessment Test, and the New York Assessment programs.

Ms. Frank is a Project Management Professional (PMP®), certified by the Project Methodology Institute (PMI®), and has an MA from the University of Northern Iowa.

Brenda Kurtz, Account Director. Brenda Kurtz (and Jeri Frank) will be PARCC's primary point of contact for contract. Ms. Kurtz has more than 14 years of program management experience at Pearson. She has managed numerous large scale state and national assessment programs, including PARCC online testing, the start-up and online launch of Readypoint Nursing, a post-secondary online testing program, the Ohio K-8 Achievement Tests, and the Hawaii State Assessments (as a subcontractor), and several postsecondary assessment programs in support of the work Pearson does for ACT.

Prior to joining Pearson, Ms. Kurtz worked for the University of Iowa Foundation and for Riverside Publishing Company. She is a certificated PMP and has a BBA in Finance from the University of Iowa.

Monica Lyons, Program Manager, Paper Forms, Packaging and Distribution, Handscoring, Psychometrics, Standard Setting, and Research. As Program Manager, Monica Lyons will manage customer service center oversight, paper forms development, packaging and distribution, handscoring, psychometrics, and research. Ms. Lyons will work closely with the other program managers at Pearson. Ms. Lyons joined Pearson in 2010, and currently serves as Program Manager for the Maryland reading/math and online assessment, New Jersey item bank, and North Carolina online summative assessment programs at Pearson.

Prior to joining Pearson, Ms. Lyons served as a coordinator for the Berry Center for Economics, Business, and Public Policy at Cornell College, and served as an in-state coordinator for a U.S. Senator. She holds a B.A. from Cornell College.

Matthew Brunscheen, **Program Manager, Item Banking and Item Development.** Mr. Brunscheen manages overall delivery of item and test development services. In his current role for the PARCC Item Development program, he develops and monitors project plans and project scope and manages overall delivery of item and test development services. Additionally, he assists with the development of training materials for item writers, content specialists, and other team members as well as facilitating customer meetings such as content review, bias and sensitivity review, data review, test construction, and forms review.

Mr. Brunscheen has 10 years of experience with Pearson and has worked on assessments for Louisiana, Illinois, Minnesota, and Puerto Rico. He is certified as a PMP and has a BA from the University of Iowa in addition to attending law school at the Universidad de Comillas Pontificia in Madrid, Spain.

Michelle Klingeman, Senior Program Manager, Operational Start Up. As Senior Program Manager, Michelle Klingeman will work closely with Ms. Frank and Ms. Kurtz as part of our PARCC Program Leadership team and will be responsible for the operational program startup activities for PARCC. Ms. Klingeman will also work closely with the Program Managers at Pearson. She will also be responsible for areas of program oversight including planning, development and execution, financial planning and monitoring, quality and customer satisfaction, and subcontractor management.

With more than 10 years of large-scale program management experience, including approximately one year on PARCC, and as a certified PMP, Ms. Klingeman brings a great deal of knowledge and understanding to the project. Her education includes an MBA from the University of Iowa.

Trent Workman, Program Manager, PearsonAccess, TestNav 8, Scoring and Reporting. Trent Workman will work with Ms. Lyons and Mr. Brunscheen to provide program management for the work associated with PearsonAccess, online testing and scoring and reporting. In his expanded role, Mr. Workman will continue to be responsible for program planning, development and execution, financial planning and functionality required for the implementation of the contract requirements. He will oversee handoffs within Pearson as well as those with PARCC. With seven years of online testing implementation and program management experience at Pearson, and approximately six months on PARCC, Mr. Workman brings a great deal of knowledge and understanding to the project.

Before joining the PARCC program team earlier this year, Mr. Workman managed the implementation of online testing across several state and national testing programs, working with both internal and external customers. In addition, he has led initiatives to streamline the creation of technology enhanced item types as well as to establish best practices for scoring and quality testing.

Mr. Workman has a BA in Business Administration in Management and Organizations from the University of Iowa and is pursuing his MBA.

Cyril Bergeron, Senior Technical Project Manager, Scoring. Cyril Bergeron is responsible for planning the scoring and reporting solution for PARCC, which includes numerous technology handoffs within Pearson. Mr. Bergeron will expand his role, planning and managing scoring and reporting for the implementation of the operational assessment. He has 11 years of experience working for Pearson in highly technical positions developing, testing, and managing software processes for state assessment contracts such as those in Minnesota and Oklahoma, and for internal Pearson products.

In addition to being certified as a PMP and being a Certified Scrum Master, Mr. Bergeron has an MBA from the University of Iowa and a BS in Engineering from the University of Madras.

Melanie Cloud Gross, Senior Project Manager, PearsonAccess. Melanie Cloud Gross will document, manage, and monitor the requirements for PearsonAccess (or for supporting the administrative portal) for PARCC and will identify and analyze issues related to the implementation of our solutions. She joined the PARCC program team in 2012 to support the Assessment Administration contract, and has been focused on the online implementation of PARCC.

Previously, Ms. Cloud Gross served as a Project Manager for Pearson's state assessment contract with Minnesota. In this role she served as a primary liaison to the Department of Education for a complex, online testing program. She collaborated with the customer to take an innovative test administration to the next level by working with internal teams to create new item types and functionality. She was also responsible for training and presented on PearsonAccess and TestNav to Minnesota.

Ms. Cloud Gross has a BA in Political Science, with a concentration in Women's Studies, from Drake University.

LeAnn Dahn, Project Manager, Call Center and Communications Management. LeAnn Dahn is responsible for managing the communications between Pearson and the participating districts and schools in PARCC and managing the call center support. Ms. Dahn most recently orchestrated 15 PARCC states, Achieve, ETS, and Pearson in recruiting a representative sample for the spring 2014 PARCC field test. This activity required taking into account the schedule, studies to be conducted, state-specific necessities, and downstream activities.

Working with PARCC states, she recruited well over 10,000 schools to participate in the field test, including the design and creation of confirmation forms as well as the delivery of state-specific test information letters to all districts.

Ms. Dahn holds a MM in Music Performance from the New England Conservatory of Music.

Kaci DeSousa, Project Manager, Paper Forms and Handscoring. Kaci DeSousa will serve as the project manager for the activities surrounding performance scoring and paper forms development. Performance scoring work includes rangefinding, coordinating aspects of training, and customer visits. Ms. DeSousa currently serves as the Pearson point of contact for PARCC and Achieve regarding the development and publishing of paper assessment materials for the PARCC spring field test.

Ms. DeSousa graduated from Wartburg College with a BA in Communications.

Tobi Drake, Senior Project Manager, Online Forms. Tobi Drake manages the items through the online forms process. Ms. Drake will serve as the primary point of contact regarding the online forms for internal Pearson departments, provide communication to PARCC states, and facilitate and document internal meetings and project review meetings with PARCC. Prior to joining the PARCC program, she worked on a contract supporting the Ohio and Hawaii programs.

Ms. Drake has a BA in Graphic Design from the University of Iowa. Additionally, she is PMP certified and holds a Teaching Certification for Art Education K–12.

Jessica Garza, Project Manager, Online Forms. Jessica Garza is responsible for managing activities associated with online forms production and technology platforms for PARCC. In addition to managing forms and technology requirements, she creates and manages project schedules encompassing cross-functional team milestones and customer hand-offs. Ms. Garza has eight years of project management experience, including two years of experience at Pearson. Prior to joining the PARCC program she provided similar services for the online Readypoint postsecondary nursing assessment program.

Ms. Garza holds a BLS from the University of Iowa.

Rebecca Gilchrist, Senior Project Manager, Schedule and Risk Management. Ms. Gilchrist develops and manages project schedules for test administration; item development coordination, forms production, printing, packing, shipping, and collection; test processing and scoring; data analysis; reporting; psychometrics, standard setting and research. She will serve as the primary point of contact regarding the project schedule for internal Pearson departments, provide communication to PARCC PARCC about the schedule, and facilitate and document internal meetings and project review meetings with PARCC.

Ms. Gilchrist is a certified PMP and a certified orange belt in Microsoft Schedules. She graduated from Mount Mercy College with a BA in Public Relations.

Jeff Heathman, Senior Project Manager, Packaging, Distribution, and Processing. Jeff Heathman will develop and manage requirements, quality processes and schedule for paper material packaging and distribution, collection of test materials for processing/scoring return, and processing of test materials including receiving, document prep, scanning, and image editing in preparation for document scoring.

As the primary point of contact regarding paper administration for all internal Pearson departments, Mr. Heathman will facilitate and document project requirements and project review meetings with PARCC.

Mr. Heathman is a PMP and has over 15 years' experience in assessment packaging and processing at Pearson, and is a current member of the PARCC program team.

Laura Kome, Senior Project Manager. Paper Form Development. Laura Kome is responsible for managing activities associated with paper forms production for PARCC. In addition to managing forms and requirements, she creates and manages project schedules encompassing cross-functional team milestones and customer hand-offs. Ms. Kome has 10 years of project management experience, including eight years of experience at Pearson. Prior to joining the PARCC program she provided project management services for our Marketing/Creative Services department as well as online and paper forms production services for the Florida program.

Ms. Kome has an MBA in Marketing and Accounting from the University of Iowa as well as an MA in Elementary Education from the University of St. Thomas.

Jennifer Tigrett, Project Manager, Online Training. For the past six years, Jennifer Tigrett has worked as a Project Manager for various technology-based projects including the PARCC/Smarter Balanced Technology Readiness Tool, ACCUPLACER, and PASeries®. She has extensive experience working internally and with customers to define, develop, and manage project plans, scope, schedule, and cost. For this contract she will serve a similar role, utilizing her vast experience with Pearson and school technology, as well as customer service.

Ms. Tigrett has worked for Pearson for 10 years and has an AAS in Computer Programming: with a Certificate in Web Design.

Lynda Valero-Garcia, Senior Project Manager, Invoicing and Subcontractor Management. Lynda Valero-Garcia will coordinate the collection of updates for status reports and the financial and invoicing record keeping. Prior to joining the PARCC program team, Ms. Valero-Garcia's most recent work at Pearson is with the Florida Department of Education on Florida Comprehensive Assessment Test (FCAT) where she managed the invoicing and the Florida Kindergarten Readiness Program delivery activities.

Holly Woodruff, Project Manager, Item Development and Translations. Holly Woodruff will serve as the primary point of contact regarding text to speech, metadata, paper, multimedia, cognitive complexity, style guide, permissions, and item bank requirements for internal Pearson departments for the PARCC program team. Ms. Woodruff has direct experience as an English Language Arts Content Specialist as a member of the PARCC item development team. Prior to joining Pearson she was the English II Program Specialist for the student assessment division of the Texas Education Agency, overseeing the development of the State of Texas Assessments of Academic Readiness.

Ms. Woodruff brings seven years of teaching experience to the program. Her BS in Radio/Television/Film is from West Texas A&M University as is her Secondary English Teacher Certification.

Pearson Content

Adrian Rivera, Test Development Manager. Mr. Rivera will oversee the development of content, including facilitating and participating in item writer training, item review meetings and other customer related meetings to create high-quality PARCC assessments. He is a Test Development Manager for the PARCC Item Development contract and will be 85% dedicated to the PARCC programs. Mr. Rivera has 13 years of experience as an editor, content assistant, content specialist and test development manager at Pearson, during which time he has developed and provided oversight for the development of assessment content for a number of state testing programs. Prior to joining Pearson he worked for the University of Texas Health Science Center. He holds an MBA and a Bachelor's of Science degree.

Martha Scarborough, Principal English Language Arts Content Specialist, ELA/Literacy Item Development, General Lead. Martha Scarborough will be the PARCC Content Lead in ELA/Literacy. She currently serves in this role for the PARCC Item Development contract. Ms. Scarborough has extensive experience managing teams of Content Specialists as they write, review, and edit items and passages for state programs such as Texas. Prior to joining Pearson, she was the Director of English Language Arts Content Development for WestEd, where she managed item development for state programs such as Pennsylvania, Nevada, and Arizona.

Her BA in English is from Harvard University; her graduate work in Curriculum and Instruction has been conducted at the University of Texas. In addition, she holds a PMP and has been a Teacher/Consultant for the National Writing Project.

Jason Rainey, Senior Content Specialist, English Language Arts Content Specialist, ELA/Literacy Item Development, Grade 3–6 Lead. Jason Rainey will be the PARCC Content Lead for grades 3–6 in ELA/Literacy. Mr. Rainey most recently served as the PARCC Content Lead for grades 9–11 in ELA/Literacy while also participating in development activities for PARCC grade 3–8.

Mr. Rainey has extensive experience writing, reviewing, and editing items and passages for state programs such as Minnesota. Prior to joining Pearson he was a middle school English teacher and a professional writer and editor.

His BA in English Literature is from the University of North Texas. Additionally, he is certified in English Language Arts for grades 6–12, a generalist for grades 4–8, and English as a Second Language (ESL) supplement.

Katherine (Kate) Brien, Senior Content Specialist, Mathematics Item Development, General Lead. Kate Brien serves as the PARCC mathematics item development general lead, participating in all aspects of mathematics test development, and conducting research to support mathematics assessment and mathematics test development. As part of the PARCC project, she communicates with the customer, is the primary PARCC mathematics contact for developing technology-enhanced tasks, coordinates high level project plans ranging from creating item development plans to creating process documents, coordinates the creation and delivery of training materials for internal team members and subcontractors, and coordinates team item development. Prior to this position, Ms. Brien was the PARCC mathematics high school lead, providing support on the development of the PARCC high school mathematics tasks.

Before joining Pearson, Ms. Brien was a middle school teacher with Austin Independent School District, responsible for teaching grade 7 magnet students, grade 8 on-level students, and developing and maintaining the scope and sequence for the grade 7 magnet mathematics program. She has also worked for the University of Texas Learning Center and KUMON.

Ms. Brien holds a BS in Mathematics as part of the UTeach program from the University of Texas.

Darren Schmidt, Senior Content Specialist, Mathematics Item Development, Grade 3–5 Lead. Darren Schmidt will serves as the PARCC mathematics lead for grades 3–8, participating in all aspects of mathematics test development, and conducting research to support mathematics assessment and mathematics test development. He works on all aspects of the test development, task development, and test construction process. Some of his responsibilities as content lead include preparing test and item specifications; writing, reviewing and editing tasks; collaborating with clients and item writers; and tracking items for test usage and selecting the appropriate items to construct test forms, including final review of all test items for both paper/pencil and online delivery.

Mr. Schmidt has experience in developing new task types for technology-enhanced tasks. He has also worked as a lead on three state projects and one shelf product. Before joining Pearson, Mr. Schmidt worked for CTB/McGraw-Hill as a mathematics content specialists working on various domestic and international projects.

Mr. Schmidt has also worked as a grade 3 and 4 teacher in the Monterey Peninsula Unified School District. He holds an MA in Education.

Pearson Knowledge Technologies

Karen Lochbaum, PhD Vice President, Technology Services. Karen Lochbaum will oversee the automated scoring components of our solution. Dr. Lochbaum holds a PhD in Computer Science from Harvard University and has been involved in the software development and management of Pearson's Intelligent Essay Assessor for almost 15 years. She holds two US patents in the development of scoring with artificial intelligence.

Don Deland, Director of Math Technology. Don Deland is responsible for Pearson's Equation Editor and associated automated math scoring. He was the president of Integre Technical Publishing prior to Pearson's acquiring the company and has been involved with automated math scoring for over 20 years.

He holds an MA in the History of Science from the University of Wisconsin—Madison and a BA in Physics and Mathematics from Oberlin College.

Pearson Software and Technology Services

Ellen Strain-Seymour, PhD, Director, Digital Management. Ellen Strain-Seymour brings research expertise from the field of human-computer interaction and whose prior research and product management experience includes the creation and study of new item types; resolving mode effects through user experience improvements; qualitative device comparability research; the interface design of embedded supports; and accessible interaction design.

Dr. Strain-Seymour manages a team of software developers, business analysts, quality assurance engineers, and user support staff for digital content systems including online item rendering, online publishing tools, item authoring, and automated item/form formatting for print delivery. She oversees staff assigned to the development of Pearson's new assessment banking and building tools specifically designed to support the most modern standards-based interoperable assessment delivery platforms (known as ABBI). The ABBI system handles item authoring, banking form building and publishing. She has eight years of online test development and content management experience at Pearson and 10 additional years of experience with e-learning planning and design, as a professor of media design, a project manager, technology researcher, and owner of software design/consulting company.

She holds a PhD in Media Research/Ethnography as well as BA and MA degrees in Communications and Design.

Ivan Horne, Manager Software Engineering. Ivan Horne manages a team of software developers responsible for the development of Pearson's proprietary item banking web application ABBI. Development responsibilities include creating an asset repository and test building software systems. He works with other teams within Pearson to manage integration between ABBI and other components such as TestNav 8 and authoring tools.

Mr. Horne has 28 years of programming, systems analyst, and management experience. He earned an AA in Applied Science–Computer Programming from San Antonio College and a BA in Music from the University of Texas at Austin.

Philip Moody, Manager, Delivery and Support. Philip Moody manages software developers, software testers, business analysts, and software support and training staff in the development of software to author, edit, and deliver online content. He provides tools and systems to expedite or automate existing digital content production processes, and is currently responsible for oversight of staff assigned to the item bank.

Mr. Moody has seven years of online content project management and software development management experience at Pearson. He has a BA in English with a minor in Education from the University of Texas.

Jason Craft, Principal Software Developer. Jason Craft is a Senior Software Developer, responsible for developing solutions for digital content, including the ABBI item and test banking system. He also researches data interoperability, HTML5 and mobile delivery, and interface accommodations for special populations.

He has three years of experience at Pearson and 10 additional years of experience as a web operations manager, technology consultant, software developer, and research fellow.

Mr. Craft earned his PhD in Digital Literacies and Literatures from the University of Texas at Austin.

Jan McSorley, Accessibility Specialist. Jan McSorley is an Accessibility Specialist for the PARCC team at Pearson. She researches accessibility solutions for test delivery systems and collaborates with the Digital Content Development team to understand the unique needs for students with disabilities. She evaluates assistive technology tools and works with internal teams to create accessible testing environments that are responsive to diverse needs. She joined Pearson earlier this year after working for the Austin Independent School District for 16 years as a special education teacher, project manager, and accessibility technology specialist.

Ms. McSorley has an MA in Curriculum and Instruction, and holds a number of accessibility certifications.

Kristy Harris, Business Analyst. Kristy Harris will be a Business Analyst for the development of the PARCC Assessment Content Repository. Ms. Harris has been in a similar position for three years and regularly collaborates with program management teams to gather, clarify, and document requirements. She will participate in review of acceptance test cases and document system capabilities and track and update changes.

Prior to her current position at Pearson, Ms. Harris held various technical writing, program management, and business analyst positions at other organizations.

She earned a BA degree in Chemical Engineering from Cornell University.

Luis Stolk, Software Quality Assurance. Luis Stolk will lead quality assurance testing for the PARCC Assessment Content Repository. He has led quality assurance for previous TestNav versions. In this role, he created and managed test cases that proved the developed software code met the requirements for the product.

Mr. Stolk has been with Pearson for two years and has six years of prior experience as a software quality assurance engineer. He holds a BS in Computer Engineering from Florida Atlantic University.

Kate Bowman, Program Technology Manager. Kate Bowman will serve as the main point of contact internally for the software team assigned to the PARCC Operational Assessments. Members of the program team or other functional groups within Pearson that have technology questions will contact Ms. Bowman. She will also communicate with PARCC regarding software requirements, issue tracking, and program fulfillment, working with the program team to coordinate and streamline all communication.

Ms. Bowman has seven years of experience at Pearson, serving primarily as a project manager for large scale assessment contracts. She has six years of prior experience with customer service and as a senior technology analyst.

Hiede Rodby, Information Technology Project Manager. Hiede Rodby will be responsible for software and testing project planning, monitoring, and technology delivery for the PARCC Operational Assessments. She will work with other team members to create the technology group's schedule, communicate project status, issues, and risks within the group, and develop the project budget. Ms. Rodby was previously the IT Project Manager for multiple assessment programs in Kentucky, Mississippi, and North Carolina.

She has 21 years of software development and IT project management experience, including 14 years at Pearson. She has a BS in Management Science from Buena Vista University.

Pat Walker, Customer Business Analyst. Patricia Walker will serve as a liaison between the Pearson program management team and the system developers, analyzing requests and requirements for development. Ms. Walker has more than 19 years of experience in business leadership as a manager, program manager, and business analyst. For the PARCC Operational Assessments, she will define the scope and deliverables based on the details of the contract. She will document requirements and track key indicators of development progress. She will be a lead member of any requirement change or process enhancement.

Ms. Walker has a BA in Business Administration, Marketing and Finance from Mount Mercy University.

Kirk Larson, Principal Developer. Kirk Larson will be a developer for the PARCC Operational Assessments program. Mr. Larson has served as lead developer of: a project to design and develop the next generation PearsonAccess application; a project to develop the Technology Readiness Tool used by both PARCC and the Smarter Balanced consortia; and the organization and student-data management modules of the PearsonAccess assessment-management system.

He has more than 20 years of experience in the design, development, testing, and implementation of software systems, including 11 in the assessment industry. His development experience includes Java, J2EE, JSP, HTML, Struts2, Struts, JSTL, SQL, MySql, Oracle, Stored Procedures, Hibernate, XML, Spring, Tomcat, BEA\Oracle WebLogic Portal, BEA\Oracle WebLogic Server. His BS in Computer Science is from Iowa State University.

Kurt Rompot, Solution Architect. Kurt Rompot will manage the overall architecture of the solution for the PARCC Operational Assessments program, Mr. Rompot has been with Pearson for more than ten years as a database administrator, product manager, and systems architect.

He has a MS in Computer Science from The University of Illinois-Springfield, and an MBA from The University of Iowa, along with numerous professional certifications.

Greg Borchert, Manager, Administration Configuration. Greg Borchert will lead the configuration of the PearsonAccess assessment management system for the PARCC

Operational Assessments program. Mr. Borchert has more than 20 years of experience as an accountant, program manager, and production manager, including over 11 years of development activity with Pearson. His familiarity with the setup and management of PearsonAccess components will enable him to coordinate configuration efforts with both Pearson and PARCC stakeholders.

He has a BA in Accounting from The University of Iowa.

Arni Storm, Manager, Performance Scoring System. Arni Storm manages development of the Pearson performance scoring system, ePEN which will be used as the handscoring system for the PARCC Operational Assessments program. Mr. Storm will continue to use his extensive knowledge of ePEN and java to oversee any scoring-system development necessary for the PARCC Operational Assessment program.

He has a BS in Electrical Engineering from The University of Iowa and more than 20 years of experience in programming.

John Gravatt, Manager, Technology Strategy. John Gravatt will help to configure the scoring systems for the PARCC Operational Assessments program. Mr. Gravatt has experience with diverse development environments and programming languages.

Mr. Gravatt holds a BS in Management Information Systems from The University of Northern Iowa.

Bill Borkowicz, Lead Developer, Packaging and Distribution System. William Borkowicz will manage technologies for the packaging and distribution system of the PARCC Operational Assessments program. Mr. Borkowicz has more than 20 years of experience as an analyst, programmer, and ITPM, including the past 15 years with Pearson. For the past 10 years, Mr. Borkowicz has worked with Pearson packaging and distribution systems.

He has a BS in Computer Science from The University of Northern Iowa.

Bryan Bleil, Vice President, Online and Technology Implementation. Mr. Bleil heads a team of implementation support specialists who provide expert guidance and training for schools, districts, and states who are new to the world of computer-based and online-delivered assessments. Mr. Bleil has also led the development and administration of the Technology Readiness Tool that the SMARTER Balanced Assessment Consortium and PARCC have both used to support states as they transition to next generation assessments. He currently leads the infrastructure trials and readiness efforts on the Assessment Administration program for PARCC, helping schools and districts prepare for computer-based testing for the spring 2014 field test.

He leads and promotes key research efforts involving use of existing and new technologies, devices, and other infrastructure (e.g., netbooks, tablets, etc.) in high-stakes statewide online testing; provides leadership for product- and process-improvement initiatives related to online testing; and collaborates with program teams, functional areas, and operations to drive improvements across the organization.

Previously at Pearson, Mr. Bleil worked as Director of Online Project for the Texas program. In that role, he directed the Texas assessment program's online testing, online training, online technology projects, and administration materials groups. He led the project and directed all efforts associated with the Texas Evaluation of Districts' Readiness for Online Testing.

Mr. Bleil holds an MS in Journalism from Columbia University with a BS in Molecular Biology from the University of Texas at Austin.

Laura Jennings, **Project Manager.** Laura Jennings, a portfolio program manager, has 15 years of experience in educational assessment, with a focus on technical process and product delivery to administrators and school staff. Ms. Jennings has spent the past three years managing the creation and delivery of training and site readiness plans for the state of Texas, which has more than 1,100 districts and more than 8,300 public schools. Ms. Jennings currently supports the Assessment Administration infrastructure training.

For this project, her responsibilities will include overseeing development of the technical readiness and training program for administration and delivery of online testing.

Ms. Jennings has a BA in Communications from the University of West Florida.

Pearson Performance Scoring

Bob Sanders, Vice President of Content, Program, and Scoring Management. Bob Sanders plans and leads scoring management support programs, including: using business analytics for planning, monitoring, and analysis of scoring operations; project delivery and support; and input into Pearson proposal activities. He provides strategic leadership to content, program, and scoring staff, including professional development and accountability expectations.

Mr. Sanders is certified as a PMP. He holds BA from the University of South Florida in Sociology, and is currently completing his MBA at the University of Iowa.

Mark Hulsebus, Program Manager, New York Scoring. Mr. Hulsebus plans, implements, and oversees training and scoring processes for several programs, including; ACT, New York, Illinois, Higher Ed, and Maryland. He coordinates the work of project managers, serving individual components of scoring projects. Mr. Hulsebus also documents and delivers customer requirements, allocating sufficient resources to required tasks and monitoring schedule, cost, and quality standards. He will be responsible for helping New York implement in-state scoring as described in Section V.B.2.K, if one of the three options is chosen.

Mr. Hulsebus holds a BS in Business Administration from Truman State University, and is a certified Project Management Professional (PMP) by the Project Management Institute (PMI). He has nearly ten years of experience in program and project management.

Amy Hauschildt, Project Manager, New York Scoring. Ms. Hauschildt will plan, implement, and oversee training and scoring processes for New York scoring. She is currently responsible for managing the overall scope and schedule of performance scoring activities, scoring ancillary creation, and for reporting status to all stakeholders throughout the lifecycle of performance scoring for the New York State 3–8 testing program as well as operational scoring activities for Aspire.
She will provide planning and oversight throughout all phases of scoring as they pertain to meeting quality standards, schedule, and contractual requirements. Ms. Hauschildt has attended rangefinding for the New York State 3–8 testing program, as well as the scoring training session in spring 2013 for New York City, gaining valuable experience, insight, and familiarity with both New York teachers and state administrators.

With more than four years' experience with Pearson, Ms. Hauschildt worked in Finance before joining Performance Scoring as a project manager. She holds an MBA from the University of Iowa and a BS in Accounting from Bellevue University.

Tamara Lyman, Scoring Program Manager. As the Program Manager for the Performance Scoring Center, Tamara Lyman will be responsible for planning, implementing, and overseeing professional human scoring required for the PARCC Operational Assessments.

Ms. Lyman has 25 years of project management experience. She is currently the Scoring Program Manager for the PARCC field test, where she will be responsible for coordinating the work or scoring project managers, content specialists, and scoring directors assigned to the project.

Her BGS in Business Administration and Management is from the University of Iowa

Margo Ballou, Scoring Project Lead. As the Scoring Project Lead for the Performance Scoring Center, Ms. Ballou will plan and implement scoring activities, including monitoring quality and schedule. A performance scoring professional, Margo Ballou has planned and directed scoring at Pearson as a project lead/content specialist, scoring director, and assistant scoring director in the subjects of reading, writing, and adult literacy and numeracy. Previously, in higher education, she taught Russian and history courses, and her research interests included textual criticism and linguistic analysis of literary texts.

Ms. Ballou's Pearson experience entails several large-scale assessment projects, including SAT, NAEP, and PARCC in the United States, and international projects such as PIAAC and PIRLS. Ms. Ballou currently serves as the Project Manager on the PARCC Field Test Administration.

In addition to PhD coursework completed in Slavic Languages at Brown University, Ms. Ballou holds an MA in Slavic Languages from the University of Texas at Austin. She is currently undertaking coursework in Project Management.

Julie Murphy, Senior Content Specialist, Mathematics. As a scoring professional, Julie Murphy supervises scoring directors, works with the Content Manager to coordinate staff assignments and balance workloads, reviews all reports and evaluations regarding the strengths and weaknesses of the scoring staff, reviews evaluations of potential scoring directors, and supervises training and development of supervisors and scorers for future content roles.

Ms. Murphy has experience as a Senior Content Specialist, Content Specialist, Content Supervisor, Project Manager, Scoring Director, Scoring Supervisor, and Scorer in math and alternate assessments. Ms. Murphy currently serves as the Senior Content Specialist on the PARCC Item Tryout and Field Test Administration.

She holds a BA in Special Education with a minor in Elementary Education from William Paterson College.

Dusti Winkie, Senior Content Specialist, ELA/Literacy. With 20 years of experience in the scoring industry, Dusti Winkie mentors scoring directors and monitors accuracy and consistency of content before and during rangefinding meetings, as well as during training development. She monitors quality and consistency during scoring across items and grades.

She has served as content specialist, project manager of content, scoring director, project manager, and scoring supervisor on assessments in a variety of subject areas, with an emphasis on ELA. She spent five years working on new projects and has assisted with multiple research studies. Ms. Winkie currently serves as the Senior Content Specialist for the PARCC Field Test Administration.

Prior to joining Pearson, Ms. Winkie taught writing and German at the secondary and post-secondary levels. She has an MA from the College of Education at the University of Iowa.

Pearson Psychometrics

Laurie Laughlin Davis, PhD, Vice President, Psychometrics. Laurie Davis has worked for Pearson for more than 10 years and is currently responsible for the strategic direction of Psychometric Services in Austin, TX, including management oversight of more than 20 research scientists and statistical analysts supporting the Texas assessment program. Her primary responsibilities include program design, development and evaluation, research/experimental design, support for government relations and communications strategies, and interpretation and communication of results of research analyses for multiple audiences. She is responsible for the on-time delivery of high quality psychometric research and operational work for the Texas assessment program.

Additionally, as part of the Research and Innovation Network, Dr. Davis leads Pearson's research efforts in the area of Digital Devices with the goal of establishing best practice for fairly assessing students on touch-screen tablet devices. In this capacity she has conducted a series of qualitative cognitive laboratories and usability studies as well as conducting quantitative research to compare student performance across computers and tablets. Dr. Davis has published white papers and has presented research at national conferences on this topic. She consults with and supports research initiatives with touch-screen tables for clients such as PARCC.

In her previous role, Dr. Davis functioned was the Director for Psychometric and Research Services. She holds PhD in Educational Psychology with a specialization in Quantitative Methods and Psychometrics from the University of Texas, and has published research in the area of computer-based testing (CBT) and computerized-adaptive testing (CAT).

Her current research interests include the incorporation of technology to enhance assessment and measurement Dr. Davis participates actively in national measurement organizations and has specific interests in mentorship and career development guidance for graduate students and new PhDs.

Her doctorate in Educational Psychology is from the University of Texas at Austin.

Ye Tong, PhD, Lead Psychometrician. Dr. Tong oversees Pearson psychometric staff and provides support to customers including PARCC, Virginia, New York, Georgia, and large districts. Additionally, she provides support for Pearson shelf products and the development of the next generation assessments. She serves as a member of the Program Leadership Team and provides overall measurement solutions for a number of state programs.

In her previous position as Manager, Psychometric Services, she was responsible for overseeing the performance of all item and test development and psychometric tasks. She has extensive operational experiences in test construction, field test and operational test designs, data analysis, equating, sampling, and standard setting. She led various teams to deliver Pearson's large state contracts, conducted research for future examination development, and worked with software developers and quality assurance analysts to verify the integrity of computerized testing systems.

Dr. Tong is a nationally renowned researcher and has published extensively in equating, vertical scaling, IRT estimation, and generalizability theory. Her most recent publications include two research articles in Educational Measurement: Issues and Practice (EM:IP) and a book chapter in Statistical Models for Test Equating, Scaling, and Linking.

Before joining Pearson, Dr. Tong was a research assistant with the University of Iowa's Iowa Testing Program. She assisted in the research and development of the Iowa Tests of Basic Skills (ITBS) and the Iowa Tests of Educational Development (ITED). She also conducted analysis, provided examples, proofread and typeset the second edition of Test Equating, Scaling and Linking by Drs. Michael Kolen and Robert Brennan.

Dr. Tong received her doctorate in Educational Measurement and Statistics from the University of Iowa.

Katie Larsen McClarty, PhD, Director, Center for College and Career Success. Katie McClarty is Director of the Center for College & Career Success in Pearson's Research & Innovation Network. She leads a team of researchers in planning and executing research in support of the Center mission, which is to identify and measure the skills needed to be successful in college and careers, determine pathways for students to be college and career ready, track their progress along the pathway, and evaluate effective ways to keep students on track.

As a lead research scientist on the development of a new statewide testing system in Texas, Dr. McClarty conducted a comprehensive research effort to align performance standards in each grade with college readiness. The research involved over 40 data studies, as well as meetings with key stakeholders such as educators from primary, secondary, and higher education, school administrators, business leaders, and legislative staff. The resulting system gives students, their parents, and teachers early indicators as to student progress toward success in college and careers after leaving high school.

Dr. McClarty has authored papers and presentations related to college readiness, standard setting, assessment design, computer-based testing, interface design, online and paper comparability, teacher effectiveness, and next generation assessments. Her work has been published in journals such as the American Psychologist, Applied Measurement in Education, Research in Higher Education, and Educational Researcher.

Dr. McClarty holds a PhD in Social and Personality Psychology from the University of Texas at Austin.

Aimee M. Boyd, PhD, Senior Research Scientist. Aimee Boyd is a Senior Research Scientist in Pearson's Psychometric Services department. Since 2011, Dr. Boyd has led the equating and scaling efforts for the Texas assessment program which includes the general and modified assessments for grades 3 through high school; Spanish assessments for grades 3–5; the linguistically accommodated assessments, and the Texas English Language Proficiency Assessment System (TELPAS) Reading for grades 3 through high school. In this role, Dr. Boyd has been responsible for leading a team of psychometricians and research analysts in the design, development, analysis, and evaluation for equating and scaling for the Texas assessment program.

Dr. Boyd led the development and implementation of new equating methodologies as Texas transitioned to a new assessment program. Her expertise in the processing and distribution of large-scale assessments has led to enhancements and efficiencies in the equating and scaling processing and verification for Texas. Dr. Boyd's role involves coordination and oversight of equating and scaling activities by an external vendor and the Texas Education Agencies' (TEA) psychometric team.

Dr. Boyd has extensive experience in the technical planning, execution, and verification for large-scale assessments through her work on the Texas project and her prior work on computer based assessments for national achievement and certification assessments at ACT.

Her doctorate in Educational Measurement is from the University of Texas at Austin.

Mike Clark, PhD, Research Scientist/Project Lead for Data Forensics. Mike Clark has led Pearson's research and development efforts in the area of data forensics since 2012. In this role, Dr. Clark has been responsible for leading a team of researchers in conducting literature reviews, designing and executing research studies, and disseminating findings.

Dr. Clark serves the measurement community as an ad hoc manuscript reviewer in the area of data forensics for several research journals, and he has presented on the topic of data forensics to a wide variety of audiences, including the Indonesian Ministry of Education, a number of state education agencies and technical advisory boards, as well as at research conferences, including NCME's annual conference, CCSSO's National Conference on Student Assessment, and the Annual Statistical Detection of Potential Test Fraud Conference.

Prior to assuming the lead role in Pearson's data forensics research and development initiative, Dr. Clark provided psychometric services for large-scale educational assessment programs in the states of Oklahoma and Florida, as well as numerous professional credentialing organizations.

His doctorate in Quantitative Psychology is from the University of Kansas.

Brian Wrobel, Manager. Brian Wrobel has supported PARCC psychometric analysis for the Research Study and Item Tryout, as well as Paper Test Map QC for the field test administration. He has seven years of experience in data analysis with Pearson and manages other Research Associates, training them on data analysis procedures and overseeing change control procedures.

In his current position, he prepares and checks critical information for score reporting, tables, and figures for statistical procedures, documentation, and reports. Additionally, he reviews results of statistical and psychometric analyses for accuracy and monitors the completion of the tasks identified in the context of the processing schedule.

Mr. Wrobel's BS in Statistics is from Iowa State University.

Pearson Quality Assurance and Continuous Improvement

Douglas Smith, Senior Quality Engineer. Mr. Smith is the point of contact for program quality. He is the Quality focal point for activities directed toward process control and quality assurance. He performs root cause analyses from quality metrics and implements corrective actions where necessary. Additionally, he identifies systemic quality issues on programs which could be implemented at the enterprise level and undertakes independent verification of project lifecycle deliverables, including consistency and adherence to quality procedures and requirements.

For over eight years, Mr. Smith has developed and implemented business process optimization strategies by analyzing existing systems, facility layout, procedures, and work flows to reduce risk, increase productivity and financial results of operations. He proposed business performance management solutions and developed detailed plans for improved methods, procedures and workflows, equipment and staffing layout, and information flows.

Mr. Smith has a BA in Management from Peru State College and continues to work on his master's degree in Organizational Effectiveness. He is also certified in Six Sigma Black Belt from the University of Alabama; Lean Manufacturing from the University of Tennessee; and ISO 9001:2000 Lead Auditor Management Rep form the British Standards Institute.

Caveon Key Personnel

Caveon Web Monitoring

Christie Zervos, Director of Operation, Web Patrol. Ms. Zervos is an integral member of Caveon's leadership team, having served as Director of Operations of Caveon's Web Patrol service for over ten years. In this role, she oversees her team's operations, maintains knowledge of internet technologies and trends, and demonstrates an unwavering dedication to her group's clients.

Ms. Zervos' team provides important web monitoring of test security threats for many large, international, high stakes test programs in all areas of testing, including education, admissions, and certification/licensure. The service detects breaches of clients' intellectual property, and provides swift resolution to mitigate any potential damage to clients' items, tests, and reputations. Her team's sterling reputation is a testament to the value the service delivers, as clients renew their service subscriptions year after year.

Prior to Caveon, Ms. Zervos worked in several leadership roles in Novell's Certification group, including overseeing test development operations. She has been an active advocate of quality testing since entering the assessment industry in 1992.

Ms. Zervos earned her BS at the University of Utah.

ETS Key Personnel

ETS Program Management

Katherine (Kit) Viator, ETS Executive Director, PARCC. Ms. Viator is the senior executive at ETS responsible for overseeing ETS' work as a subcontractor to Pearson for the Partnership for Assessment of Readiness for College and Careers (PARCC). Prior to joining ETS in 2011, she served as a senior program officer for the College-Ready Education division at the Bill & Melinda Gates Foundation where she evaluated investment opportunities and developed grants in support of the foundation's mission.

She also worked for the Massachusetts Department of Elementary and Secondary Education (MDESE), serving over 10 years as the director of student assessment, and 3 years as the state assessment coordinator. At MDESE, she held oversight responsibility for the state's K–12 academic testing programs, including the Massachusetts Comprehensive Assessment System (MCAS) program. She has served as an educational consultant to various educational interests, including the Florida Department of Education, the State of Tennessee, and the U.S. Department of Education, and has conducted educational research at Boston College and Harvard University.

She earned a MEd degree. in Administration, Planning, and Social Policy from the Harvard Graduate School of Education, and an AB degree in Humanities from the University of California at Berkeley.

Carol Owen, ETS PARCC Program Manager. Ms. Owen will be responsible for program management at ETS for PARCC and will serve as the main point of contact for within her team. From 2008 to 2012 Ms. Owen was program manager for the Computer Adaptive Achievement Test, which included developing more than 8,000 test items for the delivery of multi-stage adaptive tests to students in grades 1–11.

Between 2003 and 2008, Ms. Owen was responsible for the management of the New Jersey Assessment of Skills and Knowledge contract for students in grades 3 and 4, directing the fulfillment of the contract specifications to achieve scheduled delivery of products and services. From 1997 to 2003, she was program director for School and College Services. In that role, she directed the development of a new test series (Comprehensive Testing Program 4) used by independent schools for curriculum evaluation and student skill assessment in grades 1 through 11. Ms. Owen joined ETS in 1979 and served as associate program director and assistant examiner in the Center for Occupational and Professional Assessment.

Ms. Owen earned her MA in counseling services from Rider University and her BA in English from Wheaton College. In 2005, she completed the PMP curriculum with Northeastern University.

Sandra Wheatley, ETS PARCC Program Manager. Sandra Wheatley will offer additional program management support for the PARCC Operations Administration project and has reporting responsibilities to Kit Viator. Her cameo and resume are available upon request.

Christy Sassman, ETS PARCC Program Manager. Ms. Sassman will be responsible for ETS's delivery of the services of our statistical analysis and psychometric teams, as well as all administrative and operations aspects of our subcontract. Ms. Sassman will also be responsible for working closely with the Pearson team and PARCC representatives in support of ETS's work. She will work diligently to resolve routine and non-routine problems and any escalated issues, using sound judgment and excellent oral, written, and interpersonal communication skills.

In her role she will also inform PARCC and ETS staff about future capabilities that may enhance the program, favorably negotiate program change requests, and assess risks and proactively address potential problems to develop and maintain mutually beneficial client relationships. In addition, she will work closely with Pearson to provide program management reports and monitor program budgets, as well as conduct or participate in client meetings and team meetings.

Ms. Sassman has served as Senior Program Manager for the PARCC Item Development Contract since its award to ETS in June 2012, as well as for ETS's Assessment Administration subcontract to Pearson. Through this work, she has become deeply familiar with the PARCC program and has been very responsive to its needs.

Ms. Sassman joined ETS in 2007 as a Test Development Project Lead for K–12 assessments. Previously, she worked for Harcourt Assessment, Inc., for seven years, with her most recent role being director of educational assessment product development. She also has middle school and high school-level classroom teaching experience in mathematics.

She earned her MEd in Educational Administration and Supervision from Trinity University. She is a certified PMP.

Anthony (T.J.) Calati, Project Manager, CPMO. Mr. Calati will be responsible for ETS's delivery of the item development products and services. Mr. Calati will also be responsible for working closely with the Pearson team and PARCC representatives in support of ETS's work. He has been at ETS for more than years and has worked on a diverse number of projects, ranging from infrastructure initiatives to testing programs rollouts. Mr. Calati has a wealth of project oversight expertise in assessment development processes and practices.

He has served as PARCC Project Manager for the Item Development Contract since its award in June 2012, and most recently was promoted to Senior Program Manager for the PARCC Item Development Contract.

In another assessment projects, he managed converting a legacy of paper/pencil tests to new multi-stage adaptive tests that are delivered on-line in a new delivery platform. The project included new item development, form construction, field testing, pilots, and new score reports. He previously worked for FRABA, Inc. as the manager of the United States branch of their Posital sensor company.

Mr. Calati also holds his PMP. He earned his MS in innovation and manufacturing (A.B.T) and his BS degree in electrical engineering, both from Michigan State University.

ETS Validity Research

Brent Bridgeman, PhD, Distinguished Presidential Appointee, Validity Research. Brent Bridgeman joined ETS in 1974 after several years of college teaching. He is presently a distinguished presidential appointee in the Research & Development Division at Educational Testing Service in Princeton. Dr. Bridgeman's recent work focuses on validity and fairness issues related to test time limits, comparisons of paper-and-pencil and computer-delivered question formats, and essay assessments. He has directed or co-directed several projects since 1990 including, Effects of Extra Time on SAT® Scores, Effects of Screen Size and Resolution on Computer-Based Test Scores, and A Comparison of Open-ended and Multiple-choice Formats for the Quantitative Section of the *Graduate Record Examinations*.

Dr. Bridgeman is the recipient of fellowships from both the National Science Foundation and the NDEA. He has also published contributed to many publications, including articles in *Applied Measurement in Education* and the *Journal of Educational Measurement*. He earned his PhD in Educational Psychology from the University of Wisconsin, Madison.

Cara Cahalan Laitusis, PhD, Director, Validity Research. Cara Laitusis is the principal investigator and project director at Educational Testing Service (ETS) for three grants from the U.S. Department of Education, all of which focus on improving state assessments for students with visual impairments, blindness, learning disabilities, or mild-to-moderate cognitive impairments. She joined ETS in 1998, and her applied specializations are in curriculum-based assessment and the diagnosis and treatment of students with learning disabilities.

She has been involved in research on the validity and fairness of assessments for all test takers. These projects have included field testing of new item types for students with disabilities on both the SAT® and GRE®, examining the validity of testing accommodations for students with disabilities on a variety of tests, investigating gender differences in mathematical problem solving, and examining the comparability of paper- and computer-based test formats between gender and ethnic groups.

Dr. Laitusis has authored numerous research articles and co-edited the book *Large Scale Assessment and Accommodations: What Works?* That book was published by the Council for Exceptional Children in 2007.

She earned her PhD in Urban School Psychology from Fordham University. Dr. Laitusis will also provide consulting on accessibility issues for Students with Disabilities; Dr. Danielle

Guzman-Orth, Associate Research Scientist, will provide project oversight on accessibility for English learners.

Heather Buzick, PhD, Research Scientist. Heather Buzick works in the Center for Foundational and Validity Research at ETS. Her work at ETS involves measurement and statistical modeling for test takers with disabilities, including topics such as growth modeling, validity and fairness, mixture modeling, differential item functioning, and testing accommodations. She is currently the principal investigator and project director on a project evaluating growth measures for K12 students with disabilities funded by the U.S. Department of Education. Dr. Buzick has published in peer-reviewed journals including *Educational Researcher, Educational and Psychological Measurement,* and *Frontiers in Quantitative Psychology and Measurement* and has regularly presented her work at national conferences including AERA/NCME, the Council for Exceptional Children Convention and Expo, and the National Conference on Student Assessment sponsored by CCSSO. She earned her PhD in Measurement, Statistics, and Evaluation from the University of Maryland.

Guangming Ling, PhD, Research Scientist. Guangming Ling works in ETS's Research & Development Division. Since joining ETS in 2006 he has directed research projects related to a wide range of tests including the Major Field Tests (MFT), the National Board of Professional Teaching Standards (NBPTS), the TOEFL iBT, TOEIC, SAT, and the California Standardized Testing and Reporting (STAR).
Dr. Ling's research focuses on factors related to test validity, reliability, and fairness issues by applying latent variable models. He has authored and coauthored 24 papers in peer-reviewed research report, journals, and book chapters, including International Journal of Testing, College Students Journal, Child Development, and Educational Assessment.

He received his PhD and MA degree in Psychometrics (Quantitative Psychology) from Fordham University.

ETS Research Services

David M. Williamson, PhD, Senior Research Director. David Williamson leads the Assessment Innovations group in ETS's Research and Development Division. Dr. Williamson is responsible for establishing the vision for and tracking the execution of a research agenda targeting fundamental capability development and empirical criteria for operational implementation of automated scoring technologies.

Areas of current automated scoring capability include the automated scoring of essays for writing ability, of short responses for content, and of mathematics items and plots, as well as scoring spontaneous speech of English language learners. He manages a group of approximately 35 research scientists and engineers as they develop, evaluate and deploy new capabilities and knowledge into practice for innovative assessments for internal and external clients. Dr. Williamson was previously a research scientist with ETS from 2001 to 2006 and a senior psychometrician with the Chauncey Group International from 1996 to 2001.

He earned his PhD and MA degree in Psychometrics from Fordham University.

Danielle Guzman-Orth, PhD, ELA Researcher. Danielle Guzman-Orth will be part of the Research Advisory team for PARCC. She works in the Research and Development division at ETS on English Language Learning and Assessment (ELLA) research initiatives concerning English language assessments and English language issues related to K–12 content-area assessments in the U.S. Her research focuses on issues of appropriate assessment, identification, and accessibility for ELs and ELs with disabilities. Dr. Guzman-Orth has several manuscripts under preparation, including *The Classification Conundrum: Identifying English Language Learners At-Risk*; *Extending the Validity Framework into the Classroom: Psychometric Properties of the Working Memory Rating Scale for Spanish-speaking English Language Learners*; and *Exploring the home literacy environment for Spanish speaking English language learners at-risk for reading disabilities*.

Before joining ETS, Dr. Guzman-Orth acquired significant experience working on a longitudinal IES-funded grant to study *Growth in Literacy, Language, and Cognition for Children with Reading Difficulties Who are English Language Learners.* She also taught English language development at the K–12 level.

Dr. Guzman-Orth received both her PhD in Special Education, Disabilities, At-Risk Studies, and her master's in Special Education, Disabilities, At-Risk Studies from the University of California, Santa Barbara.

Fred Cline, Lead Research Project Manager. Fred Cline works in the Foundational and Validity Research area at ETS. Currently Mr. Cline is involved with cognitive labs to support PARCC assessment development. His primary research experience includes research design and implementation, survey construction, data analysis and graphical representation of results. Research projects Mr. Cline has been involved with primarily focus on issues of test validity and test use. His past projects include timing and speededness issues on both computer adaptive and paper and pencil assessments, and evaluating the use of automated scoring systems compared to human raters and improving the way validity data is reported. Fred also has performed an evaluation of performance assessments and instructional programs in music, reading, science, and history for K–8 students.

He earned his MBA from Rider University and his BS in Educational Research from Bucknell University.

Lauren Kotloff, PhD, Research Associate. Lauren Kotloff works in the Research & Development division at ETS. Since joining ETS in 2012, she has conducted cognitive lab studies for a variety of programs, including PARCC, NAEP Interactive Computer Task (ICT), and the Proficiency Assessment for Wyoming Students (PAWS).

Dr. Kotloff has worked at conducting cognitive interviews, analyzing data, and writing reports in assessment areas that include the Common Core State standards and students' use of scientific reasoning. Prior to join ETS she performed research for Public/Private Ventures, and published reports addressing various policy issues.

She earned her PhD in child development from Cornell University.

Elizabeth Stone, PhD, Research Supervisor. Elizabeth Stone has worked in both the statistical analysis and validity research areas since coming to ETS in 1998. In statistical analysis, her efforts concentrated primarily on coordination and statistical analysis for graduate admissions testing programs, with research focused on proficiency and parameter estimation and other practical testing issues.

Dr. Stone's work in validity research has focused on research for the Designing Accessible Reading Assessments (DARA), Technology Assisted Reading Assessment (TARA), and Feedback and Revision on AA-MAS in Mathematics grants. Other research interests include investigating fairness and validity issues for students with disabilities (particularly in the context of adaptive testing) and English language learners, issues associated with automated scoring, and other validity topics.

She earned her MS and a PhD degree in statistics from Temple University.

Teresa King, Senior Research Associate. Teresa King will provide her assessment knowledge to the PARCC program. Since coming to ETS in 2004, she has worked on projects in the Foundational and Validity Research area that examine test accessibility for students with disabilities and English language learners. Much of her work has utilized the method of cognitive labs to better understand the way test takers interact with tests and items. She has taken the lead to organize cognitive lab research studies for projects including *Graduate Record Examinations*® (GRE®), *Cognitively Based Assessment of, for, and as Learning* (*CBAL*), and has supported efforts for NAEP.

Ms. King has presented the findings from her work at regional and national conferences. She also serves as the project manager for the IES-funded research project TARA (Technology-Assisted Reading Assessment) which focuses on a program of research and development to improve reading assessments for students with visual impairments or blindness.

Ms. King earned her MA in Experimental Psychology from Long Island University.

ETS Psychometric Services

Hyeonjoo Oh, PhD, Psychometrics Manager. Hyeonjoo Oh will provide psychometric oversight for PARCC. She supervises psychometricians working on the Washington Comprehensive Assessment Program (WCAP), Proficiency Assessments for Wyoming Students (PAWS), and High School That Works (HSTW). She helps establish the technical quality of products and services, supervising, monitoring, and providing psychometric oversight and expert guidance to help effect compliance with client specifications, ETS policies, and sound measurement principles.

Previously, Dr. Oh's responsibilities included planning, designing, coordinating, and conducting statistical work for the Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT®), and she had a leading role in the SAT psychometric work. Her primary research interests are test equating, linking, and structural equation modeling. Dr. Oh has published many papers, statistical reports, and research memorandums, and given presentations in the field of psychometrics.

She earned her PhD and MA degrees in Measurement, Statistics, and Evaluation from the University of Delaware, her MA in Clinical Psychology, and her BA in Psychology from Kyungpook National University in Taegu, Korea.

Lora Monfils, PhD, Lead Psychometrician. Lora Monfils will contribute her psychometric and management knowledge to PARCC. She joined the Research & Development division at Educational Testing Service (ETS) in 2002. At ETS, Dr. Monfils has provided psychometric support as lead psychometrician, psychometric manager, or senior psychometric advisor for a wide range of K–12 state and district testing programs. She currently provides psychometric consultation services for several ETS testing programs and is leading the psychometric work in support of the PARCC field test. Dr. Monfils is the out-going former chair of the National Council on Measurement in Education (NCME) Recruitment Committee, and she has served in a number of other capacities for NCME and the American Educational Research Association (AERA).

Prior to joining ETS, she worked as a senior research analyst at the Rutgers University Center for Educational Policy Analysis. In addition, her 16 years as a classroom teacher of high school mathematics motivates her current work, including research in statistical modeling, diagnostic score reporting, and the impact of large-scale assessment on teaching practice and educational equity. Dr. Monfils is a coeditor of and a chapter author in The Ambiguity of Teaching to the Test, published by Lawrence Erlbaum Associates, Inc.

She earned her PhD in educational Psychology, Statistics, and Measurement from Rutgers University, and her BA in Fine Arts from Boston University.

Terran Brown, PhD, Senior Psychometrician. Terran Brown will oversee psychometric work on PARCC. A resume and cameo are available upon request.

Shameem Gaj, EdD, Senior Psychometrician. Shameem Gaj will provide psychometric evaluations for PARCC. Since joining ETS in 2003 she has worked on various testing programs such as SAT®, Qatar, and Middle Grades Assessment (MGA), and is currently the Lead Psychometrician for *High Schools that Work* (*HSTW*) and Washington Comprehensive Assessment Program (WCAP). Dr. Gaj assumes primary responsibility with planning, designing, coordinating, and conducting research projects and operational work for these two testing programs.

Her current research interests include changes to assessments (e.g., the impact of removing CRs from assessments) and the administration issues for online testing. Prior to joining ETS Dr. Gaj received a College Board Research Grant as a Research Fellow.

She earned her EdD degree and MA degrees in Research and Evaluation Methods at the University of Massachusetts Amherst.

Hanwook (Henry) Yoo, EdD, Psychometrician. Hanwook Yoo will provide psychometric support for PARCC. He currently works with senior staff on the California State University (CSU) English Placement Test (EPT) and Entry Level Mathematics test (ELM). Dr. Yoo was responsible for coordinating analyses and writing of technical manuals and reports for the CSU program. Prior to joining ETS he was a senior research assistant at the Center for Educational Assessment.

Dr. Yoo's research interests include applications of item-response theory (IRT) to the improvement of computer-based testing and innovative score reporting.

He earned his EdD in Research and Evaluation Methods from the University of Massachusetts at Amherst and his BSBS in English language and literature from Korea University in Seoul, Korea.

Cathy Wendler, PhD, Senior Strategic Advisor. Dr. Wendler will provide her experience and expertise to PARCC. As Senior Strategic Advisor, she provides executive consultation, providing appropriate and efficient communications between R&D and the business areas.

She is an editor for the new technical manual for the GRE® revised General Test, which details its development, relevant research, and psychometric criteria. Prior to her current role, she was Principal Director, Research Management, overseeing the Foundational and Validity Research (FVR) area in the Research & Development Division. She was also the Executive Editor for the ETS research report series.

Dr. Wendler also was the Group Executive Director for the Center for Statistical Analysis and Senior Director for assessment and research in the higher education (College Board) area in past roles. She has done research in the areas of testing accommodations and disabilities, predictive validity, gender differences, and English language learners. Dr. Wendler is currently Associate Editor for *Applied Measurement in Education*, serves on the editorial board for the AERA Division D Newsletter, and has authored a number of articles, reports, and book chapters.

Prior to coming to ETS, Dr. Wendler worked in the psychological testing field as a senior researcher, as a measurement specialist in the K-12 field, and taught courses in educational psychology and test design.

She earned her PhD in Applied Statistics and Research from the University of California, Los Angeles.

ETS Data Analysis

John Cope, ETS Data Analyst Manager. John Cope will oversee statistical analysis work for PARCC. He joined ETS in 1989 and currently provides supervision and direction for the data analyst work of five statistical associates and assistants working on the STAR, EAP, and CSU EPT/ELM testing programs. He has also supported SAT II®, *Graduate Record Examinations®* (GRE®) *Subject*, *Advanced Placement Program®* (AP®), and CAHSEE testing programs. Mr. Cope writes programs in SAS, and code for the production of reports in areas such as operational classical item analysis, DIF item analysis, IRT calibration, and IRT equating requests.

He has earned 90 credits while majoring in Electrical Engineering at Carnegie Mellon University.

Xin Xin Liu, Statistical Associate. Xin Xin Liu will provide statistical analysis support for PARCC. She is currently working on the SMARTER Balanced Assessment Consortium program as the lead data analyst, facilitating a comprehensive cross-divisional recruitment data management effort involving program management, IT, data collection services, and client teams.

Previously she led various operational data analysis tasks for the reading testing of the Washington Comprehensive Assessment Program. Ms. Liu is also a member of the statistical analysis team for the Proficiency Assessments for Wyoming Students program and lead data analysis work for mathematics testing. Before joining ETS in 2011, she worked in statistical analysis department to provide database support using SAS at a pharmaceutical company.

She earned her MS in statistics from Rutgers University and her BA in editing and publishing from Tsinghua University in Beijing.

Lin Lin, ETS Principal Statistical Associate I. Lin Lin will provide statistical support for PARCC. She joined ETS in 2004 and is currently responsible for coordinating and planning data analysis tasks for California Standards Tests. Ms. Lin has worked on several other K–12 testing programs, including the Tennessee Gateway and End-of-Course assessment the Educational Records Bureau—Computer Adaptive Achievement Test (ERB-CAAT) and the Comprehensive English Language Assessments (CELLA).
Prior to joining ETS, Ms. Lin worked as a quantitative data analyst for Rutgers University.

She earned her MS in statistics from Rutgers University, and her MA in Management—Information Science and BS in Information Science, both from Peking University, China.

Natalie Hatrak, Principal Statistical Associate. Natalie Hatrak will perform statistical analyses for PARCC. Recently she has been conducting statistical analysis and providing technical and operational support to the ELTeach program (EFT and PRK) and STAR CMA. Ms. Hatrak has also performed data analyses for all products under the TOEIC program (TOEIC LR, TOEIC SW, TOEIC Bridge, and TFI).

She began her career in 2001 as a statistical assistant with Chauncey Group International. In 2003, the TOEIC program was brought back to ETS and Ms. Hatrak moved to ETS with her other TOEIC colleagues.

She earned her MEd degree in Educational Statistics, Measurement, And Evaluation from Rutgers University and her BS degree in Mathematical Science from Penn State University.

ETS Standard Setting

Patricia Baron, EdD , Standard Setting Director and Researcher. Patricia Baron will be responsible for designing and developing materials and training facilitators in standard setting activities for PARCC. She has worked at ETS since 1987, and she brings more than 25 years of educational testing experience to the project. Since 2006, she has served as the standard setting director, researcher, and lead facilitator in our Center for Validity Research. In this role, she directs standard setting for our K–12 testing programs, including the California Standardized Testing and Reporting (STAR) program, Tennessee EOC, and Wyoming PAWS and SAWS, as well as ELL standard setting in the international context.

Dr. Baron has also designed and conducted validation studies, evaluated alignment of innovative item types to the Common Core State Standards. For the past seven years, she also has focused on research in factors contributing to variability in standard setting and development of mixed methods in curriculum and standards validity studies for state assessments. Significantly, she completed design and implementation of a standard setting tool for the Bookmark method, which provides a mechanism for expedited analysis and reporting with high quality assurance standards.

Before transitioning into her current position, she worked as the director of Government Relations and Assessment Services from 2005–2006. Previously, from 1989–2005, she was a senior psychometrician in our Research & Development division, conducting hundreds of equating and scaling studies. During her time at ETS, Dr. Baron has been the lead psychometrician on high-stakes undergraduate and graduate admissions tests, outcome assessments for college and higher-level programs, and a national assessment for Qatar. She has led development of the vertical scale and test design and helped plan standard setting for Qatar in Arabic and English. From 1992–1995, she served as assistant editor of *Educational Measurement: Issues and Practice*.

She earned her EdD in Educational Psychology, with a specialization in Educational Statistics And Measurement, from Rutgers University.

ETS Assessment Development

Nancy Glazer, PhD, Senior Director, Assessment Development. Nancy Glazer will provide senior level guidance and oversight for the PARCC program for ETS's item development work. She has worked at ETS since 1997, and in her current role, she manages ETS's Writing and Performance Assessment Group. Specifically, Dr. Glazer develops questions, scoring guides, and training materials for several writing tests, including those for the PPST® (*Praxis* Pre-Professional Skills Test), College-Level Examination Program® (CLEP®), National Assessment of Educational Progress (NAEP), and SAT®. She has extensive experience in college-level writing, having taught many writing courses (including basic composition, expository writing, and research writing) at Rutgers, Drew, and Kean Universities. She participated in the writing placement program at Rutgers, where she administered the holistically scored writing assessment to incoming students.

Dr. Glazer's areas of specialization are constructed-response design and development, reader training, and holistic scoring, and at ETS, she is currently applying this knowledge in the development of essay prompts, scoring guides, and training materials for essay and constructed-response training. She has given presentations and workshops on scoring, item development (both multiple-choice and constructed response), and the *Online Network for Evaluation*™ (ONE), and has conducted in-service training sessions for English teachers. She has presented papers at many professional conferences and is the co-author of a College Board research report, *A Survey to Evaluate the Alignment of the New SAT Writing and Critical Reasoning Sections to Curricula and Instructional Practices*.

She earned her PhD in anthropology from Rutgers University, and both her MA in Rehabilitation Counseling and a BA in anthropology from the University of Connecticut.

Todd Walker, PhD, Executive Director, Assessment Development. Todd Walker will provide executive oversight of the PARCC program. He oversees assessment development for state and district assessment programs as well as ETS product development. He manages responsibilities ranging from the hiring and training of staff to the writing, reviewing, and editing of test items. Dr. Walker joined ETS in 2002 as an item writer, and he has served in roles of increasing leadership since then. These roles have included Social Studies Assessment Specialist, Assessment Director of Educational Development, and Deputy Executive Director of K–12 Assessment Development.

During his time with ETS, Dr. Walker has developed tests and items for statewide history and social science assessments in California, Tennessee, Virginia, Oklahoma, Georgia, and Maryland as well as leading the development of a formative assessment item bank and battery of interim assessments. Before joining ETS, he taught for eight years in the Department of History at Texas Tech University. Specifically, Dr. Walker performed course design and organization for introductory American history and Western civilization courses.

Dr. Walker earned his PhD and MEd, both in History, from Texas Tech University. His extensive graduate studies have focused on US history as well as American anthropology.

Amy Johnson, Director, Assessment Development, Overall Lead Advisor. Amy Johnson will provide day-to-day supervision of the Assessment Development portion of this contract. She joined ETS in 2003 as an Assessment Specialist in English Language Arts. In her current role, she manages the PARCC coordinators and process specialist staff. Ms. Johnson has worked on a number of state assessment programs, including the California Alternate

Performance Assessment (CAPA), the California Standardized Testing and Reporting (STAR) program, and the New Jersey Assessment of Skills and Knowledge (NJ ASK), as well as ETS's internal formative assessment item bank.

Before joining ETS, Ms. Johnson worked as a language arts specialist responsible for developing district-wide instructional strategies. She also planned and presented educational workshops and training sessions based on her previous six years as a reading recovery and classroom teacher. She brings approximately 17 years of experience in the field of education to your program.

Ms. Johnson earned her MSMS in educational administration from Texas A&M University and her BA in Elementary Education from Schreiner University. In addition, she completed the Mini-MBA™ program from Rutgers University. She also holds Texas certifications in the following areas: Early Childhood Education, Elementary English 1–8, Elementary Self-Contained 1–8, and Mid-Management Administrator PK–12.

Olga Salinas, Lead Coordinator Test Development Team Lead. Olga Salinas will coordinate the scheduling, process, and work flow of PARCC item development for all high school ELA/Literacy and Mathematics and grades 6–8 Mathematics across ETS and Pearson. She began consulting with ETS in 2003 and joined the staff as a full-time mathematics assessment specialist in 2006. Two years later, Ms. Salinas was promoted to her current role of Test Development Coordinator. In this role, she develops, disseminates, and monitors item and test development schedules and workflow processes; communicates with clients and internal functional groups to help achieve quality deliverables and client satisfaction; facilitates meetings (including standard setting); and presents training sessions.

Before entering the educational testing field, she spent 13 years as a high school teacher, specializing in mathematics, science, and Spanish. Notably, Ms. Salinas was the first mathematics and science support teacher in her school district. Due to her success, every campus added this position the following school year.

She holds certifications as a secondary education teacher (mathematics and Spanish) and as a principal in the State of Texas. She earned her MA in mathematics from the University of the Incarnate Word and her BABA in Spanish from the University of Texas at San Antonio.

Carmen Dahlberg, ELA Content Lead. Carmen Dahlberg will provide content area expertise and item development oversight for PARCC. She has over 13 years of experience in assessment work. Since joining ETS, Ms. Dahlberg served as ETS collaborative co-lead on PARCC ELA/Literacy content specifications. She has worked with the team to interpret, and apply passage and item specifications, blueprints, task models, evidence statements, and PARCC item feedback. Ms. Dahlberg also taught for eight years at both the high school and middle school levels.

She earned her MEd degree in English Education from the University of Minnesota and her BA in Secondary Education Language Arts from Bethel University.

Alice Golden, English Language Arts Content Lead. Alice Golden will provide content area expertise and item development in specialized fields of reading and writing. She will establish that items meet alignment needs for the Common Core State Standards and are suitable for computer adaptive testing. Prior to joining ETS, Ms. Golden was a reading content specialist for the Minnesota Department of Education, where she provided content area expertise and item development in specialized fields of reading and writing for the Minnesota Comprehensive Assessment Series II (paper) and III (computer). Her additional positions have included senior test development specialist, educational consultant, and school teacher.

Ms. Golden earned her BS in Elementary Education from Viterbo University.

Will Steele, Overall English Language Arts Advisor, ELA Development Team Coordinator and English Language Arts Grade 10 Standard Setting Facilitator. Will Steele will provide leadership to overall English language arts development and will oversee standard setting activities on PARCC. He joined ETS in 2012 with nine years of experience in the assessment industry. Mr. Steele has served in ELA leadership and project management roles on multiple large-scale assessment programs, including the FCAT, the next generation GED, and the PARCC item development and field test forms assembly programs. He will continue to serve as an overall advisor to the ELA/Literacy team and will facilitate the standard stetting meeting for grade 10.

Mr. Steele also taught high school English for six years in New Hampshire. He earned his MA in Teaching in English and Secondary Education from the University of New Hampshire.

James Seal, English Language Arts Grade 9 Owner. James Seal will oversee activities on his grade level for PARCC. He joined ETS in 2008, and has developed passages, constructed test forms, and was the ELA team lead for the Washington Comprehensive Assessment Program (WCAP). Mr. Seal has also facilitated at item writing workshops, rangefinding meetings, and review committees for multiple programs, including PARCC 6. He has worked with ETS's guidelines committee on English language learners to develop a training plan for all content and editorial staff. Mr. Seal is also an IBIS Super User and has helped train all content areas in use of the program.

Prior to joining ETS, he worked as a middle school English teacher. He earned a BA in English from Schreiner University in Texas and a BS in Information Technology from University of Phoenix.

Justin Isenhart, PhD, English Language Arts Grade 10 Owner. Justin Isenhart will oversee activities on his grade level for PARCC. Since joining ETS in 2010, he has written, reviewed, edited, and assembled English language arts assessments for multiple state projects, including CAHSEE (California High School Exit Examination) and MS-SATP2 (Mississippi Subject Area Testing Program, Second Edition).

Prior to joining ETS, Dr. Isenhart taught at Trinity University, Connecticut College, and Harvard University. He is also a member of both the Modern Language Association and the Association of Literary Critics and Scholars.

He earned his PhD in English from Harvard University.

Michael Steier, PhD, English Language Arts Grade 11 Owner. Michael Steier will oversee activities on his grade level for PARCC. He is an Assessment Specialist II in the Higher Education Assessment division at Educational Testing Service. Dr. Steier currently manages the development of items for PARCC ELA Grade 10 and facilitates the review of PARCC materials for the PARCC collaborative Core Leadership Group as well as state educators. He also assembles and develops tests for the SAT Writing exam as well as the SAT Writing Redesign. Prior to working at ETS, Dr. Steier taught composition and literature courses at the University of Delaware.

He earned his both PhD and MA degrees in English Literature from the University of Delaware.

Shona Ruiz-Diaz, Mathematics Content Lead. Shona Ruiz-Diaz will coordinate the development of ETS across the mathematics assessments for PARCC. She arrived at ETS in 2011 with over two decades of teaching experience and works with both the Mathematics Assessment group and the Understanding Teaching Quality Center. Prior to joining our organization, Ms. Ruiz-Diaz was a teacher on special assignment for the Forest Grove School District, OR, where she wrote district-wide K–6 Mathematics assessments that were aligned to Oregon State Standards.

She earned her MA in Mathematics Education from Columbia University and her BS in Advertising and Communications from the University of Florida, Gainesville.

Theodore Slauson, Mathematics Content Lead. Theodore Slauson will oversee mathematics content on PARCC. He joined ETS in 2011 developing high stakes assessments for state clients. Previously, Mr. Slauson spent more than seven years working as an assessment specialist at both Pearson Assessment and CTB McGraw-Hill. He has more than 15 years of student instructor experience in various settings, including working for five years as the mathematics lead teacher for Everett Alvarez High School in Salinas, CA. Mr. Slauson earned his BA in mathematics and his Professional Teaching Credentials in mathematics from California State University, Sacramento. He has also attended classes on Educational Technology at the University of California, Santa Cruz Extension, and he earned a Web Application Development Diploma from MTI College.

Luis Saldivia, PhD, Overall Mathematics Advisor & Mathematics Algebra II/Integrated III Standard Setting Facilitator. Luis Saldivia will provide leadership to overall mathematics development and will support Pearson on ETS's standard setting activities for PARCC. He joined ETS in 2003 and in his most recent role serves as an Assessment Director for College

Board Programs. Dr. Saldivia was also the Director of the Educational Testing Service Research Initiative on “Test Development, Quality, Efficiency, and Innovations.” He has worked on a range of programs at ETS, including the Advanced Placement Statistics Test, the *Graduate Record Examinations*® (GRE®), and the College-Level Examination Program® (CLEP®). Prior to joining ETS, Dr. Saldivia was a college instructor and taught at both Michigan State University and the Universidad Simón Bolívar.

He earned his PhD in Mathematics from Michigan State University.

Julie Lehmann, Mathematics Development Team Coordinator. Julie Lehmann will coordinate and lead mathematics activities for PARCC. She is responsible for writing and reviewing mathematics assessments for various grade levels for multiple projects, constructing tests according to state standards, and facilitating content reviews and item-writer training meetings. Ms. Lehmann was vital in development for the following: the *Focus on Standards* (FOS) item bank project, grades 8–12; the California Standardized Testing and Reporting (STAR) program, grades 5–7; the Oklahoma Core Curriculum Tests (OCCT), grades 6–8; and the Tennessee Gateway Assessment, Algebra 1.

Before joining ETS in 2005, Ms. Lehmann was a graduate teaching and research assistant at the University of Texas at San Antonio. There, she instructed college students in a calculus lab and taught a course to review students for the Texas Academic Skills Program (TASP) test. Ms. Lehmann has also taught mathematics at the high school level.

She earned her MA degree in Mathematics Education from the University of Texas at San Antonio.

Kellie Taylor-White, Mathematics Grade 6 Owner and Mathematics Grade 6 Standard Setting Facilitator. Kellie Taylor-White, will oversee the work for her grade level and will support Pearson to oversee ETS’s standard setting activities on PARCC. She currently conducts reviews of mathematics and special education test items and forms for the Proficiency Assessment for Wyoming Students Alternate assessment program and PARCC. In addition to her work with those programs, Ms. Taylor-White has worked on the California Modified Assessment, California Alternate Performance Assessment, the Texas Assessment of Knowledge and Skills program, the Virginia Standards of Learning and Testing program, the Strengthening Instruction in Tennessee Elementary Schools—Focus on Mathematics, and the Smarter Balanced Assessment Consortium. Ms. Taylor-White also reviewed social studies test items in the field of economics for special education assessments in Florida. Before her employment with ETS, she taught middle school math and special education in both Texas and Louisiana public schools. She also served as a professional development facilitator in East Baton Rouge Parish schools.

Ms. Taylor-White is currently pursuing a Doctorate of Education in special education from North Central University, has a MA in Curriculum and Instruction from the University of Phoenix, has more than 30 hours of graduate work in the MA in Teaching—Special

Education program at Southeastern Louisiana State University, and holds a BABA in political science and sociology from Louisiana State University.

Christine P. Reyes-Swank, Mathematics Grade 7 Owner & Mathematics Grade 7-8 Standard Setting Facilitator. Christine Reyes-Swank will oversee grade 7 mathematics and support Pearson on ETS's standard setting activities for PARCC. She currently works on test designs and forms construction, and other reviews for the PARCC program. In addition to this work, she has led teams of assessment specialists working on the Washington State Measurements of Student Progress (MSP), Florida Comprehensive Assessment Tests (FCAT), Miami-Dade County Public Schools Benchmark and Interim Assessment tests, and Chicago Public Schools Benchmark Assessments. Christine has twice been the recipient of the President's Award at Educational Testing Service for her work in team building and client relations.

Prior to working in assessment, Ms. Reyes-Swank taught mathematics and self-contained classes in grades 3-6, involving herself in leadership activities and presenting professional development in several school districts in Texas.

She earned her BA in Education at Trinity University in San Antonio, TX.

Daniel Klag, Mathematics Grade 8 Owner. Daniel Klag will be responsible for coordinating the development of the items for grade 8 mathematics on PARCC. Mr. Klag will also review test forms and online forms, and he will respond to client reviews and make changes as necessary. His responsibilities have included developing mathematics items and pretest forms for the SAT®, reviewing all grade 8 mathematics items and piloting forms for PARCC, and developing new innovative item types to assess mathematics content. Mr. Klag is certified in New York to teach secondary mathematics (grades 7–12).

Prior to joining ETS he spent nearly four years as a middle school mathematics teacher. Mr. Klag has experience programming in C++ and Pascal, and he was the recipient of a research grant to investigate evolutionary computation from the Hamilton College Computer Science Department.

He earned his MA in Mathematics Education from Columbia University.

Michelle Worthington, Geometry Grade Owner. Michelle Worthington will be responsible for the geometry content on PARCC. Ms. Worthington joined ETS as a Mathematics Assessment Specialist in 2010 after working as a mathematics educator for approximately 10 years. Ms. Worthington has worked on various assessment projects at ETS, including the SAT®, the *Graduate Record Examinations®* (GRE®), *The Praxis Series*™, and the National Assessment of Educational Progress (NAEP).

Before joining ETS, Ms. Worthington taught at junior high schools and high schools in Pennsylvania and Arizona. She also worked as a teaching assistant at the University of Arizona at Tucson.

Ms. Worthington earned her MA in Teaching and Teacher Education, with a Mathematics emphasis from the University of Arizona at Tucson.

Ernest Battle, EdD, Algebra I Grade Owner & Algebra I/Integrated I Standard Setting Facilitator. Ernest Battle will oversee the work for his grade level and will support Pearson on standard setting activities on PARCC. His recent work includes writing and reviewing mathematics assessments for all grade levels across multiple projects, including reviewing mathematics test items and forms for the Tennessee EOC program. In addition to his work with Tennessee assessments, he has worked on the EQAO program in Ontario, Canada, the Texas Assessment of Knowledge and Skills program, the Tennessee Comprehensive Assessment Program and the Northwest Evaluation Association (NWEA).

Before his employment with ETS, Dr. Battle taught algebra, geometry, and advanced math in New Orleans, LA; was an assistant principal in Dallas, TX; and served as an instructional specialist in secondary mathematics for the Dallas Independent School District. He also served as a director for the Sylvan Learning Center in New Orleans.

Dr. Battle earned his EdD in Secondary Education from Indiana University.

Rocio Fletes, Algebra II Grade Owner. Rocio Fletes will oversee development for Algebra II on PARCC. She has conducted reviews of mathematics test items and forms for several programs at ETS, including the California Modified Assessment (CMA), California Standardized Test in Spanish (STS), Virginia Standards of Learning (VASOL), Formative Assessment Item Bank (FAIB), Tennessee End of Course (EOC), State of Texas Assessments of Academic Readiness (STAAR), and other programs.

Before joining ETS, Ms. Fletes worked as a graduate research assistant for faculty members in the department of economics at the University of Texas at San Antonio. She also worked as a mathematics lecturer and taught algebra and calculus to college students.

She earned her MSMS in Applied Mathematics and her MA in Economics from the University of Texas at San Antonio.

Michael Kaltman, Mathematics Integrated I, II, III. Michael Kaltman will oversee activities for his content area PARCC. Since joining ETS in June, he has worked on PARCC item development and resolution. He has also recruited for college math work and responded to item queries for the College-Level Examination Program® (CLEP®). Prior to joining ETS Mr. Kaltman taught subjects including geometry, algebra II, precalculus, basic calculus, and probability at Upper Darby School District in Drexel Hill, PA.

He earned his MEd in Secondary Mathematics from Temple University, his MS in mathematics from New York University, and his MA degree from Temple University.

Pete Flores, Integrated I, II, III. Pete Flores will oversee activities for his content area for the PARCC program. Mr. Flores works in form and item development in all phases while adhering to client specifications. Previously he was a mathematics, physics, engineering, and robotics teacher at Memorial High School and an adjunct instructor at St. Phillips College.

Mr. Flores possesses State of Texas teaching certifications in Secondary Mathematics, Physic/Physical Science, and Engineering. He received his EdD in Curriculum and Instruction from Texas A&M University, College Station, his MEd degree in Math/Science Education from Our Lady of Lake University, San Antonio; a BS in Mathematics from the University of Texas at San Antonio, and a BS in Marine Engineering from the US Naval Academy.

Kelly Van Houten-King, English Language Arts Grade 9 Standard Setting Facilitator. Kelly Van Houten-King will support Pearson and provide standard setting facilitation for PARCC. She directs ETS's staff of English Language Arts assessment specialists in writing and reviewing ELA items.

During her 11 years of work in educational testing, Ms. Van Houten-King has become adept at organizing and understanding all aspects of item development plans and blueprints for customized statewide projects, including her current work on the Maryland High School Assessment (HSA), Virginia Standards of Learning (VA SOL) tests for grade 3 through End-of-Course, and the Washington Comprehensive Assessment Program (WCAP) for grade 3 through end-of-course. Before coming to ETS in 2007, she worked for Harcourt Educational Measurement in positions of increasing responsibility, such as copy editor, senior assessment specialist, and senior product architect.

Ms. Van Houten-King also has three years of classroom teaching experience in English literature and creative writing, as well as a patent on a test development system and method. She earned her BA in English literature and French language and her MBA in a diversified major in Management, Marketing, Finance, and International Business, both from the University of Wisconsin–Madison.

Chhaya Rao, English Language Arts Grade 11 Standard Setting Facilitator. Chhaya Rao will support Pearson to provide standard setting facilitation for PARCC. She joined ETS in 2011 in a leadership position; as the PARCC ELA Content Lead for Phase I of the item development contract, she coordinated and oversaw all ELA item development and meeting facilitation across ETS and its partners.

Prior to joining ETS, Ms. Rao worked in the General Educational Development Testing Service for the American Council on Education (ACE) since 2001. Her most recent work at ACE was as Associate Director of Learning Pathways, a unit dedicated to identifying instructional pathways for GED test-takers and other adult learners. As a GED test specialist in Language Arts, she initiated an overhaul to the reading test content standards for specific literature and language benchmarks. She has also taught high school English and tutored writing at both high school and postsecondary levels. In addition, Ms. Rao has worked as a senior researcher for the National Geographic Society and a Washington, DC-based public policy consulting firm.

Ms. Rao earned her MA in English Literature from the University of Virginia and her BA in English literature and economics from Swarthmore College. She also earned her teacher certification in secondary English education from George Washington University in Washington, DC, and is a member of the National Council of Teachers of English.

Measured Progress Key Personnel

Measured Progress Program Management

Dan Verdick, Program Manager. Dan Verdick has proven his ability to establish positive working relationships across a full spectrum of stakeholders, including senior level State Department of Education clients, internal and external personnel, subcontractors, and consultants, while effectively using company resources across multiple sites. He is a strong innovator with a track record for designing strategies that exceed project expectations despite challenging constraints and changing scope.

Mr. Verdick has 15 years of experience managing and developing staff and nine years' experience managing the development, administration, and reporting of results of customized large scale statewide assessment programs. He effectively communicates design and client objectives to internal personnel and is an adept facilitator of meetings, workshops, and training for client constituencies.

Measured Progress Psychometrics

Jennifer Dunn, PhD, Director of Psychometrics. Jennifer Dunn is responsible for overseeing the daily operations of all psychometric staff, operational functions, systems development process improvements and implementing a variety of complex psychometric activities, most notably standard settings and technical reports.

Drawing on 10 years of psychometric experience, Dr. Dunn serves as the primary psychometric advisor for internal and external products, providing strategic psychometric support through the designs and implementation of research initiatives that have the potential to inform new business development. Dr. Dunn also stays current within the psychometric community by presenting at industry conferences and publishing articles.

Prior to joining Measured Progress Dr. Dunn worked at The National Center for the Improvement of Educational Assessment for three years. Since joining Measured Progress, Dr. Dunn has frequently represented Measured Progress at technical advisory meetings and has consulted on a variety of special projects both internally and externally.

In addition to her experience in psychometrics, Dr. Dunn has taught at the college level, most recently, psychometrics II at Boston College.

Louis Roussos, PhD, Senior Psychometrician. Louis Roussos joined Measured Progress in 2006. As Senior Psychometrician, he uses his extensive experience in research, analysis, and teaching to perform analysis tasks including calibrating items using Item Response Theory (IRT), equating test forms, assisting in the setting of performance standards, calculating the scaling of assessments, and collaborating with content experts to provide psychometric expertise in the design and construction of test forms.

Dr. Roussos's work includes the development of statistical procedures for conducting and supporting skills diagnosis, including standardized testing applications, data simulation, reliability, equating, and model fit. He has and continues to participate in theoretical and applied research in analyzing educational and psychological tests and related constructs; DIF analysis of educational and psychological tests; foundations and applications of nonparametric item-response-theory-based methods for detecting test multi-dimensionality, including research on DIMTEST, DETECT, and HCA/CCPROX; and theoretical and applied research on computerized adaptive testing. In conjunction with conducting this research, Dr. Roussos has extensively published technical reports, research reports, journal articles, and book chapters.

Dr. Roussos has also had more than 60 research papers presented at psychometric conferences. Dr. Roussos' success in these endeavors has been highlighted by his receiving national awards for his dissertation research (1997 NCME award, 1999 APA award) and the 2005 NCME annual award for Outstanding Example of Application of Educational Measurement Technology to a Specific Problem for his work in skills diagnosis.

Wonsuk Kim, PhD, Psychometrician. Wonsuk Kim is involved in the oversight and implementation of a variety of psychometric procedures for educational testing programs, including scaling and equating, psychometric systems development, standard setting, and other various projects.

Dr. Kim provides technical information, both verbally and in writing, to a variety of audiences. He also applies IRT models to statewide assessment systems, uses operational equating procedures, explores new methods of equating and scaling, develops and applies drifting item detection method, and participates in technical advisory committee meetings.

Seonho Shin, PhD, Psychometrician. Seonho Shin joined Measured Progress in November 2013. She has an in-depth knowledge of various theories in psychometrics and applied statistics, such as classical test theory (CTT), item response theory (IRT), generalizability theory (GT), equating, linking, and scaling. She studied computer adaptive testing (CAT), other type of automatic form assemblies, and non-parametric statistics. Dr. Shin is fluent in many commonly used statistical and measurement software programs such as SAS, R, Winsteps, PARSCALE, and BILOG-MG3 and in computer programming languages such as C++ and C#.

Prior to joining Measured Progress, Dr. Shin worked as lead psychometrician for Prometric and statistician for the University of Iowa Department of Nursing. She has provided psychometric consulting services for clients or internal staff and played a leading role in various test development projects. Her current psychometric work includes item bank management, item analysis, equating, scaling for new test development, and automated form assembly (ATA). Dr. Shin also advises internal Measured Progress staff on various issues rising from standard setting. She has extensive work experience with large scale exams such as IIM CAT in India.

WestEd Key Personnel

WestEd Test Development

Patricia Armstrong, Project Director. Patricia Armstrong serves as the Director of Test Development and the Director of Mathematics Development for WestEd's Assessment and Standards Development Services (ASDS) program. In her capacity as Director of Test Development, Ms. Armstrong directs the development of assessment services and products for general education and special populations. Ms. Armstrong led test development activities for PARCC, as well as state- and district-level projects. She led test development activities for several alternate assessments based on modified academic achievement standards, including the Pennsylvania System of School Assessments—Modified (PSSA-M) and the Kansas Assessment of Modified Measures (KAMM).

As the Director of Mathematics Development, Ms. Armstrong works closely with state departments of education to develop customized mathematics standards and item specifications to support assessment development activities.
She manages the development and selection of mathematics items for several high-stakes assessment projects, and has participated in alignment studies of state mathematics assessments.

Amy Washburn, Coordination Specialist and Project Manager. Within WestEd's ASDS program, Amy Washburn is responsible for overseeing the ASDS project coordination team. She serves as project manager on specific contracts within ASDS, and is responsible for the management and oversight of the ASDS support team that arranges staffing resource allocation within the program.

Ms. Washburn served as Project Manager for the Arizona English Language Learner Assessment (AZELLA) project. Under Ms. Washburn's management, ASDS developed the items for these assessments and produced the forms with ancillary components, including digital audio files.

Joel Carino, Content Specialist, Content Lead—Grade 7. Joel Carino serves as a Content Specialist for WestEd's ASDS program. Mr. Carino has nine years of experience in assessment development, specializes in item development for K–12 assessments, and writes original fiction and nonfiction passages. Mr. Carino led item development for assessments aligned to state and CCSS standards. He also led the development of content for an educational video game for WestEd, focusing on literacy and writing.

Mr. Carino is a published author of four nonfiction books, an author and illustrator of a children's book for the iOS platform, and has 12 years of experience as an editor and writer in nonfiction publishing.

Emily Hilligoss ,Content Specialist, Content Lead—Grade 8. Emily Hilligoss serves as a Content Specialist for WestEd's ASDS program. Ms. Hilligoss develops English language arts content for CCSS-aligned high-stakes assessments, and serves as ASDS's ELA content lead for PARCC item development for grades 6–8 and project co-lead for ELA item development for PSSA. In addition to writing and editing assessment content, she contributed to the development of ELA and English language learner standards and facilitated content review meetings.

Ms. Hilligoss holds a BA in English language and literature from the University of Chicago, and completed PhD coursework in the English department of the University of California, Berkeley.

Requirement

V.F.1.C. Communications Management
Weekly Communication Requirements
PARCC Communication—Reporting Progress
PARCC Other Communication Requirements

Response Requirements for Section V.F.1.C.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.1.C.

Deliverables for Section V.F.1.C.

a) Deliver Bi-Weekly Program Dashboard

b) Deliver Monthly Program Dashboard

c) Deliver Weekly Management Report, Meeting Notes and Action Item and Decision Log

d) Deliver Annual Program Review

Response

Complex, high-stakes assessment programs require strong leadership and coordination—and frequent communication between the customer and contractor. To ease communication for PARCC, we have designated Account Director, Jeri Frank, as the primary point of contact for the project and she will be available as a direct resource to PARCC.

Given the size of the program and the number of states involved, we recognize that a single point of contact will not be sufficient. PARCC will have full access to other members of our management and delivery team and can continue to call upon Pat Kramer, Brenda Kurtz, and John Hanson (who will be joining the team), as well as other members of our team. The approach for our communication plan is outlined in the Program Management Plan.

Communication Management and Weekly Communication

Additional components of our communication planning are as follows:

- Establish **communication protocols** in SharePoint to provide streamlined information to PARCC. This will include a communication matrix containing team members, their backups, contact information, and the required reports containing the status of milestones, accomplishments, risks/issues, and an invoicing summary.
- Provide **toll-free conference lines** and **WebEx sessions** for meetings.
- Plan and **facilitate weekly WebEx status meetings** to fully understand your requirements, review schedules, and to track progress. Agendas and meeting materials will be provided 2 days in advance of the scheduled meetings. At the conclusion of the meeting, the agenda, related materials, meeting notes, action items and decisions will be posted to SharePoint. Additionally, we will maintain a running list with outstanding action items.
- Provide a **secure method of transferring secure electronic materials**.

Reporting Progress

Reporting project progress is fundamental to successful project management. Pearson will work with PARCC to create an effective bi-weekly and monthly program dashboard that will be posted on PARCC's SharePoint site and an Annual Program Review. As needed, Pearson will participate in presentations in support of PARCC and the Partnership Manager.

Bi-Weekly Program Dashboard

Every other week, Pearson will upload a status report to the PARCC SharePoint site. This report will document the status of current and upcoming tasks, open action items, and issues requiring resolution, and high priority risks. The targeted audience for this dashboard will be the Partnership Manager.

Monthly Program Dashboard

By the last business day of each month, Pearson will upload a Monthly Program Dashboard to PARCC's SharePoint site. The report will highlight the tasks completed during the month against the milestones, open issues, high priority risks, and upcoming milestones for the following month. Pearson will keep in mind the targeted audience for this monthly dashboard will be the Partnership Manager and the State Leads, as we produce this Dashboard.

Annual Program Review

At the close of each contract year, we will produce an Annual Program Review for submission to Partnership Manager and the State Leads via PARCC's SharePoint site. Pearson understands the purpose of this this annual program review is to highlight the successes and lessons learned from the previous year. Additionally, the review will include:

- Completed milestones summary
- Issues log with root cause analyses and associated corrective actions taken
- Annotated project performance metrics
- Annotated customer satisfaction survey results
- Consolidated decision logs
- Embedded links to the specification documents.

Hard copies of the Annual Program Review will also be provided at the Comprehensive Management Meeting.

Other Communication Requirements

- During onsite visits, Pearson will provide the Partnership Manager and state staff, access to secure office space with telephones and access to the internet and printers.
- Pearson will work collaboratively with the other PARCC vendors and will adhere to the protocols laid out by the Partnership Manager on communication, document storage and calendaring.
- Pearson will provide training and access to on our schedule, risks, issues, and item authoring system.

Requirement

V.F.1.D. Scope/Change Management

Provisions Governing the Work of the Contractor

Response Requirements for Section V.F.1.D.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.1.D.

Response

Established processes verify that we systematically manage project scope and update the schedule as needed to accommodate such changes in scope. To protect project cost and time commitments, we emphasize accurate definition of the initial scope and careful control of scope changes. All PARCC program requirements will be fully documented in a Customer Requirements Allocation Document (CRAD). The program will be managed to the requirements in this document. Any scope or requirement changes will go through a change control process to carefully identify schedule, quality, and costs impacts, as well as to assess risk. Prior to implementation, changes are reviewed and approved by the program change control board (typically the Program, Technology, and Content Managers), and only after being approved are added to the CRAD.

Change Control a Shared Concern

Pearson will work with Partnership Manager to identify and discuss the potential impact of program changes on cost and schedule. We will carefully assess risks associated with schedule changes and communicate potential liability. To successfully implement a scope change for increased efficiency or effectiveness, the Partnership Manager and Pearson must agree to implement changes.

By establishing detailed requirements at the beginning, Pearson plans for the success of the entire project by noting relationships and contingencies between individual components. To drive successful outcomes, we continually assess how we are fulfilling PARCC's requirements. Our emphasis on early clarity, with detailed documentation of customer requirements, prevents unwarranted assumptions and allows our program team to track the status of each stipulation.

Requirement

V.F.1.E. Cost Management

Cost Management Requirements

Response Requirements for Section V.F.1.E.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.1.E.

Deliverables for Section V.F.1.E.

a) Deliver Quarterly Accounting Report

b) Deliver Annual Accounting Report

Response

For each quarter of contractual operation, Pearson will provide a Quarterly Accounting Report with three sections:

- Overall Annual Estimated PARCC Per-Student and Per-Test Price, Annual Estimated Aggregate Participating State Contract Value, approved change orders, executed amendments, and outstanding change orders.
- Invoice summary outlining invoices submitted to PARCC Participating States, invoices paid and outstanding compensation due to Pearson
- Expenditures

As new requirements emerge, we will collaborate with PARCC to determine the scope, price the work, and submit a change order that details the scope of the work, timeline and other relevant information in order for the Partnership Manager to approve the work to commence. We recognize the need for careful cost management and delivery of the project within budget. We will provide regular invoicing based on the contract payment schedule.

Pearson will post Quarterly and Annual Accounting Report to PARCC SharePoint site. Pearson will also be prepared to present the Annual Accounting Report at one of the Comprehensive Planning meetings.

Requirement

V.F.1.F. Quality Management

Quality Management Requirements

Response Requirements for Section V.F.1.F.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.1.F.

Deliverables for Section V.F.1.F.

a) Deliver Quality Metrics as part of the Annual Assessment Program Report

Response

Quality Management

Other organizations may talk about attention to quality, but we have taken steps to document and demonstrate the effectiveness of our quality control procedures. Pearson has adopted the ISO standard as the foundation of our quality management system. ISO 9000 offers the comprehensive quality framework necessary to demonstrate adherence to best practices and the establishment of a true culture of quality throughout our organization.

Pearson has had a quality program in place for many years which is punctuated by ISO 9001:2008 certified performance scoring, printing, distribution, scanning, and processing facilities. We embrace a philosophy of continuously improving our approach, processes and procedures to provide our students and educational partners with the highest degree of confidence in our products.

Our quality management system is defined by standardized, repeatable processes, effective customer engagement, quality assurance, quality control, and continuous improvement. Rigorous quality processes provide consistency in our products and services. Pearson quality control defines and implements critical processes at key checkpoints to verify that products and services delivered consistently meet PARCC requirements and expected quality levels.

Pearson recognizes that our ability to succeed rests on our commitment to continuous improvement. Our continuous improvement is based upon proven process engineering techniques that improve existing Pearson processes throughout the product lifecycle, with the result of providing a consistently high quality experience to our customers.

Quality Engineer

Douglas Smith, Senior Quality Engineer, will be assigned to the PARCC Operational Assessments program. The engineer is an independent member of the program core team responsible for developing and executing the program quality plan including quality assurance oversight and independent verification.

The program quality plan owned and maintained by the Quality Engineer captures the quality assurance and quality control checks for a program through the entire project lifecycle. The program quality plan is a controlled document and is maintained with other project artifacts.

Mr. Smith will be the quality focal point for activities directed towards process control and quality assurance. This includes, but is not limited to, the following activities:

- Resolves program quality issues and drives resolutions through root cause analysis and corrective actions
- Active participant in program document reviews
- Facilitates rollout of Enterprise QMS standards that may impact the program

- Independent verification of project lifecycle deliverables, including consistency and adherence to quality procedures and requirements
- Independent audit of process steps against the quality plan
- Leads quality reviews throughout the program lifecycle

Quality Management System

Program Management

We succeed at producing high quality deliverables on-time and on-budget by adhering to the principles of the Project Management Institute's (PMI) Project Management Body of Knowledge. Our business model integrates PMI processes for the sound planning, execution, and management of all our assessment projects. It is this process areas which is now undergoing certification for ISO 9001:2008 referenced above. This area of discipline includes the following:

- **Training project management staff** on effective processes in program management
- **Defining program requirements at the outset** and assessing how requirements, either original or new, affect the project plan
- **Creating a detailed project plan** for careful tracking of project deadlines and deliverables
- **Managing changes** for project effectiveness
- **Maintaining requirements-driven accountability** through to successful implementation by teaming with Organizational Quality, an independent unit within Pearson, to focus on consistent standards-driven performance and the achievement of quality goals

Using this as a foundation for our program management, we have established repeatable processes for project team coordination, and complete documentation of program processes. The features of the PMI model impart discipline and control to the Pearson project management process for predictable and consistent results. Principal features of this are manifested throughout the continuum of process areas which ultimately will meet the operational needs of the PARCC assessment as detailed below:

Software Products

Pearson uses an internal peer review process while developing software products. These formal inspections of requirements specifications, designs, source code, and other work products are a proven means for finding errors and improving the quality of software products. These reviews are conducted in accordance with documented processes.

Our software quality control procedures include the following activities:

- Preparing **detailed Customer Requirements Questionnaires (CRQs)** to identify customer requirements, identify risks, and establish priorities
- Preparing **comprehensive and detailed test plans** alongside our functional specification for a project and having key program staff sign-off on the plan
- Logging **change requests** into a database-backed workflow tracking system and assessing proposed changes for impact prior to their adoption
- Seeking **commitment from all stakeholders on all proposed change requests** and updating documents accordingly
- **Maintaining version control of all project documentation and code**
- Scheduling **testing and debugging time** in our development plan

Item and Test Creation

Pearson has proven quality assurance systems in place at all stages of the test development process, including test design, item development, field-testing, psychometric analysis, item banking, test form construction, field-test matrices development, and test booklet construction.

Our Quality Assurance processes and procedures include the following:

- **Item development and review processes** that include fact-checking, review by Pearson content specialists, and customer content and bias review committees before field-testing
- **Universal Design Review** conducted with experts with backgrounds in special education and English Language Learner instruction and assessment prior to the **Bias and Fairness Review** Committee meeting
- **Readily available replacement items** from which to draw replacements in a timely manner following internal review, if necessary
- A **well-managed test construction process** including measures for version control, enhancing validity and reliability throughout the test construction process
- **Monitoring test form alignment** with test matrices, revising to protect the viability of linking items and anchor forms, and an efficient page proof process

Scannable Forms Creation

Pearson produces scannable documents using a proven, detailed process. To maintain document quality, we work closely with the customer to design easy-to-use test forms and other materials, revising scannable documents as needed. Our iterative pre-press review process includes proofreading and customer approval for all material and media designed during each stage.

Pearson is registered ISO 9001:2008 for the quality management of printing services, meaning we have detailed procedures designed to reduce the risk of document errors. Our facilities undergo external and internal audits to verify processes are properly implemented.

Additionally, ISO 9001:2008 certification requires us to have corrective and preventive procedures in place. Should an error occur, we can demonstrate that we have corrected it and that we have taken steps to prevent it from recurring.

Additional quality control measures include the following:

- **Editorial review processes** for completeness, grammar, and layout and design
- **Sampling of every shipment** from our suppliers to confirm materials meet our standards
- **Strict printing requirements** and procedures for paper, ink, printing tolerances, and bindery activities so answer sheets scan properly

Online Test Form Production

Pearson staff in our project management and publishing operations units are highly responsive and detail-oriented. Many operational forms are approved after one just review round.

To produce quality online test forms, we use the following processes:

- **Collaboration with our customer** to clarify roles and responsibilities; refine and document review and editing processes, timelines, and communication protocols; and establish a program style guide to meet customer specifications
- **An iterative review process**, including final checks for proper form integrity and operations, culminating in customer approval
- **Weekly status calls** and a process review meeting midway through production to make any schedule adjustments and, if necessary, create an improvement plan
- **A post-project meeting** for continual improvement

Online Testing

The PARCC Assessment System allows educators to pre-load student data files, filter and sort students, view total student counts, modify student data, assign students to online and accommodated tests, manage test sessions, and view reports.

This single sign-on portal offers the following quality control advantages:

- **Enrollment management is easier** since staff closest to data also manage it, significantly reducing opportunity for errors
- Through real-time data validation and feedback, **invalid information is identified and system users are alerted** to possible key entry or transcription errors as they occur

- **Teachers save valuable time** because they can create online tests based on a group of students, a subset of students in a group, or a combination of students from multiple groups
- **Online tracking** helps you locate your shipment at every step along the way
- **Two levels of security authentication** only allow designated personnel to access sensitive data

Packaging and Distribution

Our packaging and distribution system combines industry-standard inventory management software and barcode-scanning process controls to meet the needs of large scale, high-stakes assessment programs.

For orderly, accurate, and timely packaging of test materials, our system includes the following processes:

- During **unit and product configuration testing**, Pearson's Software & Technology Systems group perform **extensive unit and product configuration testing** on distribution software programs.
- Before each administration, our Testing group **performs acceptance testing**, which includes entering enrollments and submitting student information as a mock district and processing that enrollment information through the packaging of materials in the exact manner as for live data.
- **Careful tracking of materials** as they progress from station to station for sealing, bar-coding, and shrink-wrapping prior to final shipment
- **Assigning unique barcode serial numbers to test booklets** and other secure materials for accurate packaging and shipping
- **A pick and pack process** using radio frequency scanners with a system-generated pick slip to improve accuracy
- **Automated quality control verification** to account for all materials ordered
- **Specialized handling and delivery services** and tracking capabilities from UPS
- A **pre-blue dot checklist** verifies completion of pre-production activities before blue dot production and confirms the operations departments' preparation for printing and packaging the test materials.
- The **pre-blue dot production process** uses a small sample of carefully selected districts to check for key packaging, processing, and reporting characteristics. During this process, we verify distribution materials for accuracy, completeness, print quality, and adherence to requirements.
- **Random spot-checking** occurs during packaging to verify that we adhere to specifications throughout distribution.

Receipt and Processing

Our process for the return and check in of test materials at Pearson facilities enables us to verify that all materials have been returned from schools and districts. Quality control checks for receipt and processing include the following:

- **Barcode labels** to identify and inventory returned materials
- **Identification numbers printed on each page of each student document** so individual documents can be located at any time during processing
- **Verifying machine counts of materials** against numbers indicated on school and grade identification sheets and verifying machine counts with a hand count when necessary
- **Packaging and identifying materials by batch and stack number for secure storage** so individual student documents or those of an entire district can be located within minutes

Production Control

Pearson designates a production planning analyst to work with the program management team to implement and regulate the following security procedures:

- **Reviewing your project status during each production** shift and establish guidelines to regulate flow of work, schedule personnel, and prioritize material handling
- **Assigning unique identifying numbers** to batches of test documents to carefully manage materials
- **Using a workflow management system** to track materials, obtain real-time processing information, and locate and retrieve materials

Scanning

Scanners discern between pre-printed coding and respondent markings using a 15-level mark discrimination system. The highest intensity mark is automatically selected as the intended mark. Other quality control measures for Pearson scanning processes include the following:

- During **unit testing**, we review and approve unit test plans before executing tests, to confirm that software components are complete before we begin product configuration testing.
- During **product configuration testing,** we use multiple test decks (both manual and automated) to verify proper collection and editing of data for answer documents.
- To **establish editing rules** we create a separate test deck, using approved documented test cases standardized across administrations. This process also improves accuracy by reducing issues, and reduces time required for this verification.
- **Acceptance testing** consists of a **mock set of answer documents** to verify our scanning program captures marks accurately and processes them correctly

- **Checking the output file** against each answer document after processing
- **A formal sign-off practice** to verify that processing activities are performed correctly
- **Labeling each document** with a unique identification number
- **Scanning technology that alerts operators** if a scanner requires service
- A **pre-blue dot checklist** verifies completion of pre-production activities before the blue dot process, and confirms that the operations departments are prepared for processing live materials.
- The **blue dot** consists of specified districts containing pre-determined criteria needed to provide data capture and editing quality. During this activity, the Testing group verifies proper capture and editing of data from unique scannable documents, based on a pre-defined sampling of materials. We do this to provide accurate live processing and to confirm adherence to customer requirements.

Selected Response Scoring

Pearson's integrated development system for designing, printing, and proofing our own forms results in efficient, accurate scoring of documents.

Optical mark reader scanners capture student data from scannable forms used in the assessment. Electronic data files created from this process are validated in the following stages:

- **System test of scan processes**
- **Integrated test of the scan-to-score and score-to-report processes, mirroring the production environment**
- **Customer review and approval of the output**

Performance Scoring

All Pearson scorers have earned college degrees, and all receive identical training on individual test items. Our standard system helps us maintain quality and consistency in large-scale handscoring projects.

The advanced functionality of our image-based scoring system offers the following advantages:

- **Real-time, on-demand access** to comprehensive statistical monitoring reports
- Calibration sets that provide examples of scoring criteria and situations for **consistent scoring across challenging items**
- **An automated, diagnostic back reading tool** to monitor scorers, identify issues, and provide constructive feedback
- **A flexible, transparent validity system** to check accuracy across scorers

Results Analysis

To verify valid and accurate data analysis and reporting, Pearson uses the following processes:

- **Requirements Documentation** for collaboration with the customer to completely document your contract's analysis and reporting requirements so all involved will have a clear understanding of necessary tasks
- **Decision Rules** to establish criteria for issues documentation, reports distribution, and aggregation of out-of-district students
- **Specifications Conformance** to verify that all expected data fields are in the data file and that each field is populated with valid values
- **Item Documentation** when checking for item keys and reporting codes distribution by subject area
- **Key Check for Item Analysis** for verifying answer keys by examining psychometric performance of items, checking items that have a high degree of non-response, and statistical examination of multiple-choice items to identify potentially mis-keyed items
- **Check Electronic Scoring** to confirm the number of students scored per test form and the value of raw scores and score distributions
- **Complete Data Analyses** to check data analysis and psychometric programs comply with decision rules and results of data analyses are accurate and complete
- **Report Files and Score Reports** to verify the numbers of schools/students, value of scores, standard errors, ranges, weighted means of scores, and samples of reports at each aggregate level

Score Reports

Both the customer and Pearson will verify that reports meet requirements.

Our software development team determines score report accuracy, including identifying any discrepancies between expected and actual results. To maintain accuracy, Pearson processes include:

- During **unit testing**, we review and approve unit test plans before test execution, to confirm the completeness of software components before product configuration testing.
- During **product configuration testing**, we use multiple test decks (both manual and automated) to verify the accurate referencing of scoring keys and the proper application of aggregation rules according to requirements. We independently test every unique score point for all unique scoring combinations. In addition, we create separate test decks to test the aggregation and matching combinations. We review test cases for completeness before test execution. Records generated through this process also contain expected results, which we then compare to the system's actual output.
-

This extensive test provides accurate scoring and adherence to requirements.

- **Acceptance testing** consists of emulating receipt of live materials for proper processing. Test material is processed by the operations area that will be processing the live material during production. The Testing group examines output to verify that data are scored, aggregated, and matched properly. During this process, the same materials used to verify the materials distribution and data collection systems are used to verify the scoring system. This process also verifies that test data are processed in the same manner as live material.
- **Processes for testing** reports and listings, production run books, report assembly, and mailing
- **Comparison of report formats** to input sources from approved customer samples
- **Evaluation of production run book performance** by comparing to customer requirements
- **A first production batch test** will validate a subset of systems with examples of key reporting circumstances representative of the whole
- **Customer-selected school divisions** incorporated and the last check prior to mailing reports is performed

Reports Production

The following Pearson quality control measures verify that reports are packaged and distributed correctly to districts and schools.

- **Reports assembled based on project specifications** and packing lists or customer-defined and documented project specifications
- **A banner page** segregating reports into appropriate units
- **A print quality check** completed to confirm that reports are readable and select data is present
- **Verification that all reports** are present and have been packaged and labeled correctly
- **Error records maintained** to determine trends and areas where improvement is needed
- We perform **pre-blue dot** verification to confirm the completion of pre-production activities before the blue dot process and to confirm the operations departments' preparation for generating the production materials. This process entails generating all reporting deliverables in the production environment.
- In the **blue dot process,** based on specific demographic criteria, we select a sample of districts awaiting reports. We produce reports in the production environment as if they would be sent to the districts. After packaging of pre-blue dot reports, the Testing staff verifies them for accuracy and completeness and adherence to requirements.
- **We randomly sample reports** during packaging to confirm adherence to specifications throughout the distribution.

Quality Auditing

Staff from Organizational Quality, an independent unit within Pearson, conducts the following internal audits of quality control procedures to verify effective implementation.

- **An audit schedule** developed and maintained to adhere to Pearson's established processes
- **An audit plan created, the audit conducted and the findings reported** to process owners and relevant Pearson management
- **Audit nonconformities resolved** through the audit corrective action process, which involves problem definition, corrective action, cause analysis, and verification of action effectiveness

The audit team has access to all management levels and functional areas.

Continual Improvement

Pearson recognizes that our ability to succeed rests on our commitment to continual improvement. Our continual improvement is based upon proven process engineering techniques that improve existing Pearson processes throughout the product lifecycle, with the result of providing a consistently high quality experience to our customers.
Our Continual Improvement model is based on W. Edward Deming's well-known and proven Plan-Do-Check-Act cycle:

- **Plan.** Identify an opportunity and plan for change.
- **Do.** Implement the change on a small scale.
- **Check.** Use data to analyze the results of the change and determine whether it made a difference.
- **Act.** If the change was successful, implement it on a wider scale and assess results. If the change did not work, begin the cycle again.

Requirement

V.F.1.G. Risk Management

Response Requirements for Section V.F.1.G.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.1.G.

Deliverables for Section V.F.1.G.

a) Deliver Monthly Risk Register

Response

We define a **risk** as an event or problem that could potentially occur at some point in the future. Although outside the control of the project, it could affect the project's objectives, quality, budget, or schedule.

Pearson uses risk management as a means of proactively identifying and mitigating risk events to decrease the probability and effect of threats to the program.

Risk identification remains an iterative process. While risks will be identified early in the project, identification continues throughout the duration of the contract. Risks are documented and controlled in a centralized location and the Risk Register will be provided to the state leads and Partnership Manager once a month on 5th working day of each month through the PARCC SharePoint site. Risk response planning involves developing strategies on how to address risks of low, moderate, and high priority and the mitigation strategy to reduce the threat of the risk, therefore delivering on-time and in a manner that meets PARCC expectations. The risk register will describes the risk, the owner of the risk, the probability and impact of a realized risk, risk impact rating, overall risk rating, and recommended mitigation and contingency plans.

Requirement

V.F.1.H. Schedule Management

Response Requirements for Section V.F.1.H.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.1.H.

b) Offeror's proposal shall include a draft project schedule for the first operational year.

Deliverables for Section V.F.1.H.

a) Deliver Monthly Program Schedule.

Response

At the beginning of each test administration year, Pearson will develop a comprehensive project schedule of the activities associated with sections V.A – V.F of the RFP. Our master scheduler, Rebecca Gilchrist, will coordinate with our internal teams and subcontractors to develop the schedule.

The schedule will reflect the project activities, activity durations, hand-offs, and dependencies for the project. The schedule is an ever living breathing document. Ms. Gilchrist will update the schedule weekly and once a month will post to PARCC's SharePoint site. The schedule will be available both in PDF format and Microsoft Project. In the event there are significant changes between the monthly submissions, Pearson will provide an updated schedule.

A draft schedule is included in the **Other Supporting Materials** section of the proposal. The schedule captures many critical milestones associated with the first year of the program; however additional time and input (including that from PARCC) will be required to further develop the schedule. We anticipate the schedule will go through multiple versions before being baselined.

V.F.2 Partnership Manager Roles and Responsibilities

Requirement

V.F.2.A. The Offeror shall provide assurances that the Offeror, including its SubContractors, will work collaboratively with the Partnership Manager and other PARCC Vendors.

Response Requirements for Section V.F.2.A.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.2.A

Response

Good project management and good working relationships with our customers are a key element of the success of any program, particularly on a very complex startup program like PARCC. Pearson will work through the Partnership Manager, who has the primary authority and responsibility for the work under this contract.

The Partnership Manager will assign the appropriately skilled staff to oversee the Operational Assessment contract in areas of test development; administration, including scoring and reporting psychometric services; and technology-based services. Likewise, Pearson has assigned highly-skilled leads across these management areas. The Partnership Manager will also have responsibility for coordinating the work with the PARCC states. These program leads will work closely with the Partnership Manager on schedules, specifications, design of products, scope, and payment schedules and will provide have five working days to review and approve products.

V.F.3 Program Management Meetings

Requirement

V.F.3.A. Program Management Meetings

Response

Pearson will identify meeting locations, coordinate the meeting logistics (including electronic participation), and be responsible for the costs of the meeting facilities. Additionally, Pearson will be responsible for the travel costs of the relevant Pearson team members, our subcontractors, and the Partnership Manager and Partnership Representative travel.

We are planning for two meetings each year for the following areas:

- Comprehensive program management
- Test development
- Assessment administration

- Psychometric services
- Standard setting (year 1 only)

Additionally, we have provided costs for up to seven meetings each year to cover project specific meetings. Topics for these meetings may include the following:

- Handscoring
- Technology
- AAF
- Content Development

Once a meeting date and high-level agenda have been established, Pearson will provide a meeting invitation and logistical materials to all potential participants. We will also provide a contact name and number to handle their specific questions prior to and following the meeting. We will provide the invitation at least one month before the meeting. The final agenda and documented outcomes will be approved one week prior to the meeting. Following approval, Pearson will prepare electronic and paper copies of all relevant materials for the meeting.

At the conclusion of each meeting, we will post the agenda, participant lists, meeting notes, action items, and decisions to the SharePoint site. We will maintain an ongoing action and decision log for each component of the contract.

Requirement

V.F.3.B. Partnership Representative Travel to Management Meetings

Response

The costs associated with travel for the Partnership Manager and other relevant participants are included in the Per-Student and Per-Test pricing established in the Contract. Pearson will reimburse participants for their travel to the meetings listed below and provide a quarterly and annual accounting of the expenditures from this fund.

Partnership Representative Travel to Management Meetings						
Meeting #	**Meeting**	**Meetings/ Year**	**Days/ Meeting**	**Number of PARCC Representatives**	**Likely Location**	**Approx. Cost per Meeting**
1	Comprehensive Program Management Meeting	2	2	15	Contractor Site	$24,000
2	Project Specific Meetings	7	4	10	Contractor Site	$22,000
3	Test Development Planning Meetings	2	2	12	Contractor Site	$19,200

Partnership Representative Travel to Management Meetings						
Meeting #	**Meeting**	**Meetings/ Year**	**Days/ Meeting**	**Number of PARCC Representatives**	**Likely Location**	**Approx. Cost per Meeting**
4	Assessment Administration Planning Meetings	2	2	12	Contractor Site	$19,200
5	Psychometric Services Planning Meetings	2	2	12	Contractor Site	$19,200
6	Standard Setting Planning Meetings	2 in year 1 only	1	15	Contractor Site	$19,700

Requirement

V.F.3.C. PARCC Oversight Meetings

1. Partnership Manager Responsibilities
2. Contractor Responsibilities

Response

Each year, there will be eight state lead meetings and three Technical Advisory Committee (TAC) meetings in Washington, DC as listed in the following figure:

State Lead and Technical Advisory Committee Meetings							
Meeting #	**Meeting**	**Meetings/ Year**	**Days/ Meeting**	**Number of Travelers**	**Number of Attendees/ Meeting**	**Likely Location**	**Approx. Cost/ Meeting**
1	State Lead	8	3	20	30	Washington, DC	$50,000
2	TAC	3	1	25	35	Washington, DC	$45,000

Meetings in Washington, DC. Each year there will be eleven meetings in Washington, DC for which Pearson is provides some funds.

For these meetings, Pearson will be responsible for the following items:

- Hotel logistics (hotel room, meeting space and audio/visual equipment)
- Coordinating meal options
- Coordinating travel for participants
- Providing clear, timely communication to participants
- On-site organization and support (participant registration, meeting support and point of contact for the hotel)
- Meeting materials (participant badges and tents and meeting signage)

Pearson understands the Partnership Manager will have responsibility for the following items:

- Providing the initial participant list
- Developing the meeting agenda
- Developing the meeting materials
- Facilitating the meeting
- Documenting meeting notes, decisions and action items

Pearson will send the appropriate staff members to the meetings for logistical support and to best support the topics of discussion on the agenda. Pearson will be responsible for our own staff and subcontractor travel. At the conclusion of the meeting, Pearson will manage the travel reimbursements for the participants. These travel costs will come out of the PARCC Oversight Meeting Fund. On a quarterly and annual basis, Pearson will provide a detailed accounting of the expenditures from this fund.

Requirement

Response Requirements for Section V.F.3.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.3.

Deliverables for Section V.F.3.

a) Deliver Meeting Agendas, Participant Lists, Meeting Materials, Meeting Notes (excluding notes from oversight meetings) and Action Item and Decision Log (excluding action items and decisions from oversight meetings)

Response

Once a meeting date and high level agenda have been established, Pearson will provide participants with meeting invitations, logistical materials, and a contact person to handle questions. This will be provided at least one month prior to the meeting. The final agenda and documented outcomes will be approved one week prior to the meeting. Following approval, Pearson will prepare electronic and paper copies of relevant materials for the meeting.

At the conclusion of each meeting, the agenda, participant lists, meeting notes (with the exception of notes from the oversight meetings), action items, and decisions will be posted to the SharePoint site (with the exception of action items and decisions from oversight meetings).

V.F.5 Final Delivery of Materials

Requirement

V.F.5. A. The Contractor agrees to deliver to the Partnership Manager all materials and products in all forms that are developed for and used in conjunction with this program, including test books, answer documents (PDF), final electronic source files of interpretive products, software source code, and ancillary materials, and electronic files within 30 days following acceptance by the Partnership Manager of the final report for the program. Payment of the final program invoice will not be made until all

materials and certification of destruction, as appropriate, are received and approved by the Partnership Manager and final payment resolution is agreed to by both parties.

Response Requirements for Section V.F.5.

a) Offeror's proposal shall include a response to the requirements specified in Section V.F.5.

Response

When the project is completed, Pearson will deliver to the Partnership Manager all copies of the materials and products (all forms in all formats) produced throughout the lifetime of the contract. We will provide these within 30 days of the final report acceptance. We acknowledge the final invoice for the program will not be paid until the Partnership Manager has received the final certification of destruction of materials and there is resolution on the amount of the final payment by both entities.

V.F.6. PARCC Data Privacy Guidelines

Requirement

V.F.6.A. The Contractor shall comply with the Partnership's Data Privacy Guidelines. The Data Privacy Guidelines are pending final approval and will be posted to parcconline.org on or about November 20, 2013.

Response

Pearson will comply with the consortium-wide policy Data Privacy Guidelines approved on December 5, 2013, at the Governing Board meeting. With the adoption of this policy, PARCC states have set stringent policies and procedures that allow states and districts to control their data, maintain the security of data, and strictly limit access to data for only specific purposes that support the state assessment program. Pearson acknowledges the Data Privacy and Security Policy prohibits the sharing of student data with any outside entity, including the federal government, without explicit direction from individual states.

V.G. Cost Saving Recommendations

Not Applicable

Contract Terms and Conditions

Requirement

15. Contract Terms and Conditions

The contract between an agency and a contractor will follow the format specified by the Agency and contain the terms and conditions set forth in the Sample Contract (APPENDIX D). However, the contracting agency reserves the right to negotiate provisions in addition to those contained in this RFP (Sample Contract) with any Offeror. The contents of this RFP, as revised and/or supplemented, and the successful Offeror's proposal will be incorporated into and become part of any resultant contract.

The Agency discourages exceptions from the contract terms and conditions as set forth in the RFP Sample Contract. Such exceptions may cause a proposal to be rejected as nonresponsive when, in the sole judgment of the Agency (and its evaluation team), the proposal appears to be conditioned on the exception, or correction of what is deemed to be a deficiency, or an unacceptable exception is proposed which would require a substantial proposal rewrite to correct.

Should an Offeror object to any of the terms and conditions as set forth in the RFP Sample Contract (APPENDIX D) strongly enough to propose alternate terms and conditions in spite of the above, the Offeror must propose specific alternative language. The Agency may or may not accept the alternative language. General references to the Offeror's terms and conditions or attempts at complete substitutions of the Sample Contract are not acceptable to the Agency and will result in disqualification of the Offeror's proposal.

Offerors must provide a brief discussion of the purpose and impact, if any, of each proposed change followed by the specific proposed alternate wording.

Response

Pearson agrees to the Sample Contract terms and conditions included in the RFP as Appendix D and does not take any specific exceptions to these terms and conditions. However, to the extent that certain terms are not applicable, Pearson presumes that those terms will not be included in the final negotiated contract as provided in Question #81 in the document, Vendor Questions and Responses. By way of example, Pearson is not offering in this Proposal the licensing of any Pearson Intellectual Property and we presume that the provisions in the Sample Contract related to licensing and source code will not be included in the final contract as they are not applicable. Any Pre-Existing Pearson Intellectual Property and any derivatives thereof that are used to deliver the services outlined in this proposal will be owned by Pearson, and the Agency will not receive any on-going rights to this Intellectual Property.

PARCC

Offeror’s Terms and Conditions

Requirement

16. Offeror's Terms and Conditions

Offerors must submit with the proposal a complete set of any additional terms and conditions they expect to have included in a contract negotiated with the Agency.

Response

Pearson does not have any additional terms and conditions to be considered in a contract negotiated with the Agency.

PARCC

Plan to Mitigate Conflict of Interest/ Affidavit

Pearson recognizes that currently held contracts for PARCC Item Development (State of Florida Contract #12-685) and Assessment Administration (with PARCC Inc) may constitute an organizational conflict of interest (OCI) for conducting some of the activities included in the PARCC Operational Assessment RFP (40-000-13-00027). We also understand that the perception of OCI could erode confidence in PARCC's assessments, even if no conflict truly exists. For those reasons, we will subcontract with Measured Progress to provide quality control for psychometrics, as described in this proposal. Measured Progress is a full-service assessment provider; however, their scope of work for the proposed contract is limited to psychometric quality control.

Furthermore, we understand that our responsibilities on current PARCC contracts cannot unduly favor the methodologies, outcomes, and deliverables associated with this RFP to the benefit of Pearson, however per the response to PARCC Operational Assessment RFP Vendor Questions and Responses #10, *"Florida awarded contracts to vendors for development work. They are not incumbent vendors for purposes of an operational assessment contract awarded by New Mexico, and they will be eligible to submit proposals in this solicitation. Many states have a long history of releasing solicitations for assessment services and allowing the incumbent vendor to submit a proposal in response to the solicitation. Relevant prior experience is not an OCI situation and does not constitute an unfair competitive advantage requiring mitigation or avoidance."*

In addition to assigning some of the work to Measured Progress, we will also subcontract work to Caveon, ETS, and WestEd as shown in the table below.

Division of Responsibilities	
Component	**Primary Responsibility**
Test Development	ETS, Pearson, WestEd
Assessment Administration	Pearson
Psychometric Services	ETS, Pearson Measured Progress (Quality Control) Caveon (Data Forensics: Internet Monitoring)

Division of Responsibilities	
Component	**Primary Responsibility**
Reporting	Pearson
Standard Setting	Pearson ETS, WestEd (Content Facilitators)
Program Management	Pearson

PARCC Operational (Rev 4/7/14)

Meetings Details (Year 1: Feb 2014-June 2015)

Meeting Details									
Meeting Type/(Cost Reporting Category)	In-Person Meetings Per Year	Virtual Meetings Per Year	Meeting Duration Per Meeting	# of State Participants Per Meeting	# of PARCC Staff Attending Per Meeting	Total Travelors Per Year (State+PARCC)	Number Participants Eligible for Stipends/Substitutes Per Year	Location
Text Review (content development)	2	0	5	76	6	164	72	Hub City
Core Leadership Group (CLG) Review (content developmen	4	0	5	61	6	268	0	Hub City
State Educator (SE) Review (content development)	2	2	4	76	6	164	288	Hub City
Bias/Sensitivity (B/S) Review (content development)	2	0	4	52	6	116	192	Hub City
Editorial Review (content development)	2	0	5	13	2	30	0	Hub City
Technical APIP Review (content development)	0	0	0	0	0	0	0	N/A
Data Review (content development)	0	0	0	0	0	0	0	N/A
Common Form Pulling Review (forms construction)	2	0	3	20	4	48	0	Contractor Site
Matrix Form Pulling Review (forms construction)	0	2	5	18	3	0	0	Virtual
Form Review (forms construction)	0	2	3	18	3	0	0	Virtual
Rangefinding Review (scoring)	2	0	5	90	6	192	180	Hub City
Rangefinder Review (scoring)	2	0	5	18	6	48	0	Washington DC
Test Deck (quality assurance)	3	0	5	8	2	30	0	Contractor Site
HS Monitoring (scoring)	3	0	10	2	7	27	0	Contractor Site
Rule-based Scoring (scoring)	0	1	5	4	4	0	0	Virtual
Cut Score Review/Calibration (psychometrics)	0	0	0	0	0	0	0	Contractor Site
Standard Setting (psychometrics)	0	0	0	0	0	0	0	Hub City
Program Management Meeting (PM)	4	0	2	6	9	60	0	Contractor Site
Project Specific Meetings (PM)	10	0	4	5	5	100	0	Contractor Site
Test Development Planning Meetings (PM)	3	0	2	4	8	36	0	Contractor Site
Assessment Administration Planning Meetings (PM)	2	0	2	6	6	24	0	Contractor Site
Psychometric Services Planning Meetings (PM)	3	0	2	8	4	36	0	Contractor Site
Standard Setting Planning Meetings (PM)	3	0	1	8	7	30	0	Contractor Site
State Lead (PM)	12	0	3	20	10	360	0	Washington DC
TAC (PM)	4	0	1	25	10	140	0	Washington DC
Other: CLG Item Review Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Contractor/PARCC Reconciliation Debrief	0	1	1	30	5	0	0	Virtual
Other: State Educator Reconciliation Meeting	0	1	1	30	5	0	30	Virtual
Other: State Educator Reconciliation Debrief	0	1	1	30	5	0	30	Virtual
Other: Bias/Sensitivity Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Bias/Sensitivity Reconciliation Debrief	0	1	1	30	5	0	0	Virtual
Other: Editorial Review Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: PLD Webinars	0	6	4 hours	49	5	0	49	Virtual
Other: Item Development Strategy Meeting	0	2	1	30	5	0	0	Virtual
Other: Review of Psychometric Analyses	1	0	2	2	4	6	0	Hub City
Other: Field Trial of Standard Setting	0	0	0	0	0	0	0	N/A
Other - Describe								
Other - Describe								

Note: PM = Program Management

Enter Standard Stipend Rate $150.00 Substitue Payment $ 100.00

Travel Assumptions Should use GSA rates for all per diems - Enter Override rate if applicable **Hub city: $195/night, DC: $225/night**

Data in red represents changes from original proposal per final contract negotiations and price reduction. Stipend rate of $150/day or substitute rate of $100/day was applied as directed on pages 69-72 of the RFP. TAC stipends of $1500/day were also applied as noted on RFP page 173. Meeting locations noted as "hub city" are assumed to be held in PARCC hub cities with GSA rates comparable to that of Chicago, Illinois. Please note, although "Technical APIP Review" is listed as a meeting in the template and in Appendix U, third party review by a subcontractor and a face-to-face meeting are n required.Data Review, Cut Score Review, and Standard Setting are shown in the template but do not occur until year 2. The Field Trial of Standard Setting was removed from the requirements during negotiations.

Meetings Details (Year 2: Feb 2015-June 2016)

Meeting Details

Meeting Type/(Cost Reporting Category)	In-Person Meetings Per Year	Virtual Meetings Per Year	Meeting Duration Per Meeting	# of State Participants Per Meeting	# of PARCC Staff Attending Per Meeting	Total Travelors Per Year (State+PARCC)	Number Participants Eligible for Stipends/Substitutes Per Year	Location
Text Review (content development)	1	1	5	76	6	82	72	Hub City
Core Leadership Group (CLG) Review (content development)	4	0	5	61	6	268	0	Hub City
State Educator (SE) Review (content development)	4	0	4	76	6	328	288	Hub City
Bias/Sensitivity (B/S) Review (content development)	2	0	4	52	6	116	192	Hub City
Editorial Review (content development)	1	1	5	13	2	15	0	Hub City
Technical APIP Review (content development)	0	0	0	0	0	0	0	N/A
Data Review (content development)	1	0	5	118	12	130	60	Hub City
Common Form Pulling Review (forms construction)	2	0	3	20	4	48	0	Contractor Site
Matrix Form Pulling Review (forms construction)	0	2	5	18	3	0	0	Virtual
Form Review (forms construction)	0	2	3	18	3	0	0	Virtual
Rangefinding Review (scoring)	4	0	5	90	6	384	360	Hub City
Rangefinder Review (scoring)	2	0	5	18	6	48	0	Contractor Site
Test Deck (quality assurance)	3	0	5	8	2	30	0	Contractor Site
HS Monitoring (scoring)	3	0	10	2	7	27	0	Contractor Site
Rule-based Scoring (scoring)	0	1	5	4	4	0	0	Virtual
Cut Score Review/Calibration (psychometrics)	1	0	3	3	3	6	0	Contractor Site
Standard Setting (psychometrics)	3	0	4	*	*	273	249	Hub City
Program Management Meeting (PM)	2	0	2	6	9	30	0	Contractor Site
Project Specific Meetings (PM)	7	0	4	5	5	70	0	Contractor Site
Test Development Planning Meetings (PM)	2	0	2	4	8	24	0	Contractor Site
Assessment Administration Planning Meetings (PM)	2	0	2	6	6	24	0	Contractor Site
Psychometric Services Planning Meetings (PM)	2	0	2	8	4	24	0	Contractor Site
Standard Setting Planning Meetings (PM)	1	0	1	8	7	15	0	Contractor Site
State Lead (PM)	8	0	3	20	10	240	0	Washington DC
TAC (PM)	3	0	1	25	10	105	0	Washington DC
Other: CLG Item Review Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Contractor/PARCC Reconciliation Debrief	0	1	1	30	5	0	0	Virtual
Other: State Educator Reconciliation Meeting	0	1	1	30	5	0	30	Virtual
Other: State Educator Reconciliation Debrief	0	1	1	30	5	0	30	Virtual
Other: Bias/Sensitivity Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Bias/Sensitivity Reconciliation Debrief	0	1	1	30	5	0	0	Virtual
Other: Editorial Review Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Item Development Strategy Meeting	0	2	1	30	5	0	0	Virtual
Other: Review of Psychometric Analyses	1	0	2	2	4	6	0	Hub City
Other - Describe								
Other - Describe								

Note: PM = Program Management

Enter Standard Stipend Rate → $150.00 Substitue Payment $ 100.00

Travel Assumptions Should use GSA rates for all per diems - Enter Override rate if applicable **Hub city: $195/night, DC: $225/night**

Data in red represents changes from original proposal per final contract negotiations and price reduction. Stipend rate of $150/day or substitute rate of $100/day was applied as directed on pages 69-72 of the RFP. TAC stipends of $1500/day were also applied as noted on RFP page 173. Meeting locations noted as "hub city" are assumed to be held in PARCC hub cities with GSA rates comparable to that of Chicago, Illinois. Please note, although "Technical APIP Review" is listed as a meeting in the template and in Appendix U, third party review by a subcontractor and a face-to-face meeting are not required.

PARCC Operational (Rev 4/7/14)

Meetings Details (Year 3: Feb 2016-June 2017)

Meeting Details								
Meeting Type/(Cost Reporting Category)	In-Person Meetings Per Year	Virtual Meetings Per Year	Meeting Duration Per Meeting	# of State Participants Per Meeting	# of PARCC Staff Attending Per Meeting	Total Travelors Per Year (State+PARCC)	Number Participants Eligible for Stipends/Substitutes Per Year	Location
Text Review (content development)	1	1	5	76	6	82	72	Hub City
Core Leadership Group (CLG) Review (content developme	2	2	5	61	6	134	0	Hub City
State Educator (SE) Review (content development)	2	2	4	76	6	164	288	Hub City
Bias/Sensitivity (B/S) Review (content development)	2	2	4	52	6	116	192	Hub City
Editorial Review (content development)	0	2	5	13	2	0	0	Hub City
Technical APIP Review (content development)	0	0	0	0	0	0	0	N/A
Data Review (content development)	1	0	5	118	12	130	60	Hub City
Common Form Pulling Review (forms construction)	2	0	3	20	4	48	0	Contractor Site
Matrix Form Pulling Review (forms construction)	0	2	5	18	3	0	0	Virtual
Form Review (forms construction)	0	2	3	18	3	0	0	Virtual
Rangefinding Review (scoring)	4	0	5	90	6	384	360	Hub City
Rangefinder Review (scoring)	2	0	5	18	6	48	0	Contractor Site
Test Deck (quality assurance)	3	0	5	8	2	30	0	Contractor Site
HS Monitoring (scoring)	3	0	10	2	7	27	0	Contractor Site
Rule-based Scoring (scoring)	0	1	5	4	4	0	0	Virtual
Cut Score Review/Calibration (psychometrics)	0	0	0	0	0	0	0	N/A
Standard Setting (psychometrics)	0	0	0	0	0	0	0	N/A
Program Management Meeting (PM)	2	0	2	6	9	30	0	Contractor Site
Project Specific Meetings (PM)	7	0	4	5	5	70	0	Contractor Site
Test Development Planning Meetings (PM)	2	0	2	4	8	24	0	Contractor Site
Assessment Administration Planning Meetings (PM)	2	0	2	6	6	24	0	Contractor Site
Psychometric Services Planning Meetings (PM)	2	0	2	8	4	24	0	Contractor Site
Standard Setting Planning Meetings (PM)	0	0	0	0	0	0	0	N/A
State Lead (PM)	8	0	3	20	10	240	0	Washington DC
TAC (PM)	3	0	1	25	10	105	0	Washington DC
Other: CLG Item Review Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Contractor/PARCC Reconciliation Debrief	0	1	1	30	5	0	0	Virtual
Other: State Educator Reconciliation Meeting	0	1	1	30	5	0	30	Virtual
Other: State Educator Reconciliation Debrief	0	1	1	30	5	0	30	Virtual
Other: Bias/Sensitivity Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Bias/Sensitivity Reconciliation Debrief	0	1	1	30	5	0	0	Virtual
Other: Editorial Review Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Item Development Strategy Meeting	0	2	1	30	5	0	0	Virtual
Other: Review of Psychometric Analyses	1	0	2	2	4	6	0	Hub City
Other - Describe								
Other - Describe								

Note: PM = Program Management

Enter Standard Stipend Rate $150.00 Substitue Payment $ 100.00

Travel Assumptions Should use GSA rates for all per diems - Enter Override rate if applicable **Hub city: $195/night, DC: $225/night**

Data in red represents changes from original proposal per final contract negotiations and price reduction. Stipend rate of $150/day or substitute rate of $100/day was applied as directed on pages 69-72 of the RFP. TAC stipends of $1500/day were also applied as noted on RFP page 173. Meeting locations noted as "hub city" are assumed to be held in PARCC hub cities with GSA rates comparable to that of Chicago, Illinois. Please note, although "Technical APIP Review" is listed as a meeting in the template and in Appendix U, third party review by a subcontractor and a face-to-face meeting are not required.

PARCC Operational (Rev 4/7/14)

Meetings Details (Year 4: Feb 2017-June 2018)

Meeting Details

Meeting Type/(Cost Reporting Category)	In-Person Meetings Per Year	Virtual Meetings Per Year	Meeting Duration Per Meeting	# of State Participants Per Meeting	# of PARCC Staff Attending Per Meeting	Total Travelors Per Year (State+PARCC)	Number Participants Eligible for Stipends/Substitutes Per Year	Location
Text Review (content development)	0	1	5	76	6	0	72	Hub City
Core Leadership Group (CLG) Review (content developme	0	2	5	61	6	0	0	Hub City
State Educator (SE) Review (content development)	2	2	4	76	6	164	288	Hub City
Bias/Sensitivity (B/S) Review (content development)	2	2	4	52	6	116	192	Hub City
Editorial Review (content development)	1	1	5	13	2	15	0	Hub City
Technical APIP Review (content development)	0	0	0	0	0	0	0	N/A
Data Review (content development)	1	0	5	118	12	130	60	Hub City
Common Form Pulling Review (forms construction)	2	0	3	20	4	48	0	Contractor Site
Matrix Form Pulling Review (forms construction)	0	2	5	18	3	0	0	Virtual
Form Review (forms construction)	0	2	3	18	3	0	0	Virtual
Rangefinding Review (scoring)	2	0	5	90	6	192	180	Hub City
Rangefinder Review (scoring)	2	0	5	18	6	48	0	Contractor Site
Test Deck (quality assurance)	3	0	5	8	2	30	0	Contractor Site
HS Monitoring (scoring)	3	0	10	2	7	27	0	Contractor Site
Rule-based Scoring (scoring)	0	1	5	4	4	0	0	Virtual
Cut Score Review/Calibration (psychometrics)	0	0	0	0	0	0	0	N/A
Standard Setting (psychometrics)	0	0	0	0	0	0	0	N/A
Program Management Meeting (PM)	2	0	2	6	9	30	0	Contractor Site
Project Specific Meetings (PM)	7	0	4	5	5	70	0	Contractor Site
Test Development Planning Meetings (PM)	2	0	2	4	8	24	0	Contractor Site
Assessment Administration Planning Meetings (PM)	2	0	2	6	6	24	0	Contractor Site
Psychometric Services Planning Meetings (PM)	2	0	2	8	4	24	0	Contractor Site
Standard Setting Planning Meetings (PM)	0	0	0	0	0	0	0	N/A
State Lead (PM)	8	0	3	20	10	240	0	Washington DC
TAC (PM)	3	0	1	25	10	105	0	Washington DC
Other: CLG Item Review Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Contractor/PARCC Reconciliation Debrief	0	1	1	30	5	0	0	Virtual
Other: State Educator Reconciliation Meeting	0	1	1	30	5	0	30	Virtual
Other: State Educator Reconciliation Debrief	0	1	1	30	5	0	30	Virtual
Other: Bias/Sensitivity Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Bias/Sensitivity Reconciliation Debrief	0	1	1	30	5	0	0	Virtual
Other: Editorial Review Reconciliation Meeting	0	1	1	30	5	0	0	Virtual
Other: Item Development Strategy Meeting	0	2	1	30	5	0	0	Virtual
Other: Review of Psychometric Analyses	1	0	2	2	4	6	0	Hub City
Other - Describe								
Other - Describe								

Note: PM = Program Management

Enter Standard Stipend Rate $150.00 Substitue Payment $ 100.00

Travel Assumptions Should use GSA rates for all per diems - Enter Override rate if applicable **Hub city: $195/night, DC: $225/night**

Data in red represents changes from original proposal per final contract negotiations and price reduction. Stipend rate of $150/day or substitute rate of $100/day was applied as directed on pages 69-72 of the RFP. TAC stipends of $1500/day were also applied as noted on RFP page 173. Meeting locations noted as "hub city" are assumed to be held in PARCC hub cities with GSA rates comparable to that of Chicago, Illinois. Please note, although "Technical APIP Review" is listed as a meeting in the template and in Appendix U, third party review by a subcontractor and a face-to-face meeting are not required. Some meetings occur only in Year 1 & 2 (e.g., standard setting) as noted above. Rangefinding in summer 2018 and fall 2018 for the items field tested in the spring of 2018 are in support of future operational forms and costs are not included as part of this contract.

Y1 Winter 2014-15 HS PBA - 100% PPT

2,291 Schools
1,896 Distribution Points
1,656 **Systems/Districts**

	Document Title	Document Description	# Pgs	# of Fms	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
Scannable Materials			84	9	0	0	525,000	525,000	500,000		
ELA--PBA	Grade 9 ELA -- Answer Book	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	12	1	S	Y	93,399	93,399	88,952	20's, 5's - or other TBD	
	Grade 10 ELA -- Answer Book	" "	12	1	S	Y	88,226	88,226	84,025	"	
	Grade 11 ELA -- Answer Book	" "	12	1	S	Y	80,875	80,875	77,024	"	
MATH--PBA	Algebra I Answer Book	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	16	2	S	Y	93,403	93,403	88,955	"	
	Algebra II Answer Book	" "	16	2	S	Y	88,234	88,234	84,033	"	
	Geometry Answer Book	" "	16	2	S	Y	80,863	80,863	77,012	"	
Scannable Headers			2	1			0	15,119	13,745		
	School/Classroom Header	scannable sheet, 166 (6x6 layout), 9.5x11, mark reflex, black plus one additional color, 60# offset, continuous, lithocode front/back	2	1	S	Y	16,631	15,119	13,745	Custom	Placed in District and/or School Coordinator kit
Non-scannable Test Books/Scripts			264	9	0	0	525,000	525,000	0		
ELA--PBA	Grade 9 ELA --Test Booklet	8-3/8X10-7/8 booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color + black on cover, one color + black internal	48	1	NS	Y	93,399	93,399		Custom to order	Custom to secure order assignment
	Grade 10 ELA --Test Booklet	" "	48	1	NS	Y	88,226	88,226		"	
	Grade 11 ELA -- Test Booklet	" "	48	1	NS	Y	80,875	80,875		"	
MATH--PBA	Algebra I Test Booklet	" "	40	2	NS	Y	93,403	93,403		"	
	Algebra II Test Booklet	" "	40	2	NS	Y	88,234	88,234		"	
	Geometry Test Booklet	" "	40	2	NS	Y	80,863	80,863		"	
Large Print			264	6	0	0	1,943	1,943	0		
	Large Print Grade 9 ELA Test Book	large print book, 14X17, 150% photo enlargement, black on white 50# stock, coil bind (minimum 18-pt type)	48	1	NS	Y	345	345		1's, Kit	Kits to include instruction memo, corresponding form 1 test book & matching scan A/doc
	Large Print Grade 10 ELA Test Book	" "	48	1	NS	Y	327	327		"	" "
	Large Print Grade 11 ELA Test Book	" "	48	1	NS	Y	299	299		"	" "
	Large Print Algebra I Test Book	" "	40	1	NS	Y	346	346		"	" "
	Large Print Algebra II Test Book	" "	40	1	NS	Y	326	326		"	" "
	Large Print Geometry Test Book	" "	40	1	NS	Y	299	299		"	" "
Braille			660	6			191	185	0		
	Braille Grade 9 ELA Test Book	Contracted Braille: 11-1/2x11" Manila 100# stock	120	1	NS	Y	33	33		1's, Kit	Kits to include instructions and matching scan A/doc
	Braille Grade 10 ELA Test Book	" "	120	1	NS	Y	31	31		"	" "
	Braille Grade 11 ELA Test Book	" "	120	1	NS	Y	29	29		"	" "
	Braille Algebra I Test Book	" "	100	1	NS	Y	33	33		"	" "
	Braille Print Algebra II Test Book	" "	100	1	NS	Y	34	31		"	" "
	Braille Print Geometry Test Book	" "	100	1	NS	Y	31	28		"	" "
Other Non-Scannable Printed Materials (Manuals, etc.)			384	3			26,342	23,947	0		
	Test Coordinator Manual (Paper)	8-1/2 x 11 self cover booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	96	1	NS	Y	4,342	3,947		1's	Placed in District and School Coord Kits
	Test Admin Manual (Paper) Gr 3-5 ELA	8-1/2 x 11 self cover booklet, perfect-bound, 50# white offset, prints one color plus black on cover, black all interior			NS	Y	0	0		5's, 1's	

	Test Admin Manual (Paper) - Gr 9-11 ELA	8-1/2 x 11 self cover booklet, perfect-bound, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	144	1	NS	Y	11,000	10,000		TBD - 5's & 1's or other	
	Test Admin Manual (Paper) Gr 9-11 Mathematics	" "	144	1	NS	Y	11,000	10,000		"	
Printed & Other Non-Stock Materials (CD. Cassettes, Etc.)			18	11			541,408	541,408	0		
	Algebra I Reference sheet	8.5 x 11 Single sheet, printed black simplex on 50# white offset, Wrap in 20's	2	1	NS	Y	98,073	98,073		20's or TBD	Will come from the vendor wrapped in 20's or other TBD size
	Algebra II Reference sheet	" "	2	1	NS	Y	92,646	92,646		"	" "
	Geometry Reference sheet	" "	2	1	NS	Y	84,906	84,906		"	" "
BR	BRAILLE - Algebra I Reference sheet	Contracted Braille: 11-1/2x11" sheet, Manila 100# stock	2	1	NS	Y	33	33		1's, Kit	Part of Braille kit
BR	BRAILLE - Algebra II Reference sheet	" "	2	1	NS	Y	31	31		"	" "
BR	BRAILLE - Geometry Reference sheet	" "	2	1	NS	Y	28	28		"	" "
	Ruler - Grade - 6-HS	Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	NS	N	262,500	262,500		TBD	
LP	LG Print - Ruler - Grade - 6-HS	LP version to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	NS	N	971	971		1's, Kit	Part of Lg Print kit
BR	BRAILLE - Ruler - Grade - 6-HS	Brailled to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	NS	N	92	92		1's, Kit	Part of Braille kit
	Large Print Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	NS	N	1,943	1,943		"	Insert in Lg Print kit assemby
	Braille Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	NS	N	185	185		"	Insert in Braille kit assemby
	Read Aloud Kit	Provide uncollated (test book) units for use with students with special accommodations to have the test read aloud in small groups.						0			Nothing known related to any additonal materials.
Miscellaneous Materials (Stock)			3	2			606,095	652,193	0		
	Pre-ID Labels	LABEL, PRE-ID, 3900, 3.5X1 7/16 TONER BACK	0.2	1	NS	Y	525,000	525,000			
	Paper Bands	standard 3"x27" brown kraft self-sealing bands	1	1	NS	N	35,000	35,000			
	Stock Colored Return labels	3x5 colored return labels, 2-up, laser compatible	0.5		NS	N	15,365	15,365			
	Stock Outbound Shipping labels	3x5 white shipping labels, 2-up, laser compatible	0.5		NS	N	15,365	15,365			
	UPS Labels	As needed for carrier. Stock	0.5		NS	N	15,365	15,365			

Y1 Winter 14 HS EOY

Y1 2014-15 Winter HS EOY - 100% PPT

2,291 Schools
1,896 Distribution Points
1,656 **Systems/Districts**

	Document Title	Document Description	# Pgs	# of Fms	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
Scannable Materials			42	6	-	-	525,000	525,000	500,000		
ELA--EOY	Grade 3 ELA -- Consumable Test Booklet	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode			S	Y	-	-		20's, 5's 80:20 ratio	
	Grade 9 ELA -- Answer Doc	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	2	1	S	Y	93,399	93,399	88,952	20's, 5's - or other TBD	
	Grade 10 ELA -- Answer Doc	" "	2	1	S	Y	88,226	88,226	84,025	"	
	Grade 11 ELA -- Answer Doc	" "	2	1	S	Y	80,875	80,875	77,024	"	
MATH--EOY	Algebra I Answer Book	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	12	1	S	Y	93,403	93,403	88,955	"	
	Algebra II Answer Book	" "	12	1	S	Y	88,234	88,234	84,033	"	
	Geometry Answer Book	" "	12	1	S	Y	80,863	80,863	77,012	"	
Scannable Headers			2	1			16,631	15,119	13,745		
	School/Classroom Header	scannable sheet, 166 (6x6 layout), 9.5x11, mark reflex, black plus one additional color, 60# offset, continuous, lithocode front/back	2	1	S	Y	16,631	15,119	13,745	Custom	Placed in District and/or School Coordinator kit
Non-scannable Test Books/Scripts			240	21	-	-	525,000	525,000	-		
ELA--EOY	Grade 9 ELA --Test Booklet	8-3/8X10-7/8 booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson standard is 50# white offset), prints one color + black on cover, one color + black internal	48	5	NS	Y	93,399	93,399		Custom to order	Custom to secure order assignment
	Grade 10 ELA --Test Booklet	" "	48	5	NS	Y	88,226	88,226		"	
	Grade 11 ELA -- Test Booklet	" "	48	5	NS	Y	80,875	80,875		"	
MATH--EOY	Algebra I Test Booklet	" "	32	2	NS	Y	93,403	93,403		"	
	Algebra II Test Booklet	" "	32	2	NS	Y	88,234	88,234		"	
	Geometry Test Booklet	" "	32	2	NS	Y	80,863	80,863		"	
Large Print			240	6	-	-	1,943	1,943	-		
	Large Print Grade 9 ELA Test Book	large print book, 14X17, 150% photo enlargement, black on white 50# stock, coil bind (minimum 18-pt type)	48	1	NS	Y	345	345		1's, Kit	Kits to include instruction memo, corresponding form 1 test book & matching scan A/doc
	Large Print Grade 10 ELA Test Book	" "	48	1	NS	Y	327	327		"	" "
	Large Print Grade 11 ELA Test Book	" "	48	1	NS	Y	299	299		"	" "
	Large Print Algebra I Test Book	" "	32	1	NS	Y	346	346		"	" "
	Large Print Algebra II Test Book	" "	32	1	NS	Y	326	326		"	" "
	Large Print Geometry Test Book	" "	32	1	NS	Y	299	299		"	" "
Braille			600	6			191	185	-		
	Braille Grade 9 ELA Test Book	Contracted Braille: 11-1/2x11" Manila 100# stock	120	1	NS	Y	33	33		1's, Kit	Kits to include instructions and matching scan A/doc
	Braille Grade 10 ELA Test Book	" "	120	1	NS	Y	31	31		"	" "
	Braille Grade 11 ELA Test Book	" "	120	1	NS	Y	29	29		"	" "
	Braille Algebra I Test Book	" "	80	1	NS	Y	33	33		"	" "
	Braille Print Algebra II Test Book	" "	80	1	NS	Y	34	31		"	" "
	Braille Print Geometry Test Book	" "	80	1	NS	Y	31	28		"	" "
Other Non-Scannable Printed Materials (Manuals, etc.)			-	-	-	-	-	-	-		
Printed & Other Non-Stock Materials (CD. Cassettes, Etc.)			26	14			553,948	553,948	-		
	Algebra I Reference sheet	8.5 x 11 Single sheet, printed black simplex on 50# white offset, Wrap in 20's	2	1	NS	Y	98,073	98,073		20's or TBD	Will come from the vendor wrapped in 20's or other TBD size
	Algebra II Reference sheet	" "	2	1	NS	Y	92,646	92,646		"	" "
	Geometry Reference sheet	" "	2	1	NS	Y	84,906	84,906		"	" "

LP	Lg Print - Algebra I Reference sheet	Single Lg Print 14X17 sheet, 150% photo enlargement, black on white 50# stock (minimum 18-pt type)	2	1	NS		346	346		1's, Kit	Part of Lg Print kit
LP	Lg Print - Algebra II Reference sheet	" "	2	1	NS		326	326		"	" "
LP	Lg Print - Geometry Reference sheet	" "	2	1	NS		299	299		"	" "
BR	BRAILLE - Algebra I Reference sheet	Contracted Braille: 11-1/2x11" sheet, Manila 100# stock	2	1	NS	Y	33	33		1's, Kit	Part of Braille kit
BR	BRAILLE - Algebra II Reference sheet	" "	2	1	NS	Y	31	31		"	" "
BR	BRAILLE - Geometry Reference sheet	" "	2	1	NS	Y	28	28		"	" "
LP	LG Print - Ruler - Grade 4-5	LP version to support: Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock			NS	N	93,403	93,403		"	" "
LP	LG Print - Ruler - Grade - 6-HS	LP version to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	NS	N	181,637	181,637		1's, Kit	Part of Lg Print kit
BR	BRAILLE - Ruler - Grade - 6-HS	Brailled to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	NS	N	92	92		1's, Kit	Part of Braille kit
	(All) Large Print Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	NS	N	1,943	1,943		"	Insert in Lg Print kit assemby
	Braille Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	NS	N	185	185		"	Insert in Braille kit assemby
	Read Aloud Kit	Provide uncollated (test book) units for use with students with special accommodations to have the test read aloud in small groups.						-			Nothing known related to any additonal materials.
Miscellaneous Materials (Stock)			3	2			593,924	627,848	-		
	Pre-ID Labels	LABEL, PRE-ID, 3900, 3.5X1 7/16 TONER BACK	0.2	1	NS	Y	525,000	525,000			
	Paper Bands	standard 3"x27" brown kraft self-sealing bands	1	1	NS	N	35,000	35,000			
	Stock Colored Return labels	3x5 colored return labels, 2-up, laser compatible	0.5		NS	N	11,308	11,308			
	Stock Outbound Shipping labels	3x5 white shipping labels, 2-up, laser compatible	0.5		NS	N	11,308	11,308			
	UPS Labels	As needed for carrier. Stock	0.5		NS	N	11,308	11,308			

Y1-Y4 Traditional Spr 3-11 PBA

12,460 Schools
5,647 Distribution Points
2,252 Systems/Districts

	Document Title	Document Description	# Pgs	# of Fms	Y2	Y3	Y4	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
Scannable Materials			544	242	229	146	144	-	-	6,825,697	6,825,697	6,500,000		
ELA--PBA	Grade 3 ELA -- Consumable Test Booklet	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	56	6	6	6	4	S	Y	493,923	493,923	470,403	Custom to order need	Custom to secure order assignment
	Grade 4 ELA -- Answer Book	" "	12	3	3	3	3	S	Y	498,538	498,538	474,798	20's, 5's - or other TBD	
	Grade 5 ELA -- Answer Book	" "	12	3	3	3	3	S	Y	496,505	496,505	472,862	"	
	Grade 6 ELA -- Answer Book	" "	12	3	3	3	3	S	Y	492,246	492,246	468,806	"	
	Grade 7 ELA -- Answer Book	" "	12	3	3	3	3	S	Y	492,937	492,937	469,464	"	
	Grade 8 ELA -- Answer Book	" "	12	3	3	3	3	S	Y	486,968	486,968	463,779	"	
	Grade 9 ELA -- Answer Book	" "	12	3	3	3	3	S	Y	160,614	160,614	152,966	"	
	Grade 10 ELA -- Answer Book	" "	12	3	3	3	3	S	Y	151,709	151,709	144,485	"	
	Grade 11 ELA -- Answer Book	" "	12	3	3	3	3	S	Y	139,059	139,059	132,437	"	
MATH--PBA	Grade 3 Math Consumable Test Booklet	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	40	25	25	15	15	S	Y	493,218	493,218	469,731	Custom to order	Custom to secure order assignment
	Grade 4 Math Answer Book	" "	12	15	15	9	9	S	Y	497,826	497,826	474,120	20's, 5's - or other TBD	
	Grade 5 Math Answer Book	" "	12	15	15	9	9	S	Y	495,795	495,795	472,186	"	
	Grade 6 Math Answer Book	" "	12	15	15	9	9	S	Y	491,543	491,543	468,136	"	
	Grade 7 Math Answer Book	" "	12	15	15	9	9	S	Y	492,233	492,233	468,793	"	
	Grade 8 Math Answer Book	" "	12	15	15	9	9	S	Y	486,272	486,272	463,116	"	
	Algebra I Answer Book	" "	16	15	15	9	9	S	Y	120,284	120,284	114,556	"	
	Algebra II Answer Book	" "	16	20	15	11	11	S	Y	113,611	113,611	108,201	"	
	Geometry Answer Book	" "	16	15	15	9	9	S	Y	104,155	104,155	99,195	"	
	Int I Answer Book	" "	16	15	13	5	5	S	Y	40,094	40,094	38,185	"	
	Int II Answer Book	" "	16	15	13	5	5	S	Y	37,876	37,876	36,072	"	
	Int III Answer Book	" "	16	20	16	5	5	S	Y	34,718	34,718	33,065	"	
	Spanish Gr 3 Math Consumable Test Booklet	" "	40	1	1	1	1	S	Y	806	806	672		
	Spanish Gr 4 Math Answer Book	" "	12	1	1	1	1	S	Y	814	814	678		
	Spanish Gr 5 Math Answer Book	" "	12	1	1	1	1	S	Y	811	811	676		
	Spanish Gr 6 Math Answer Book	" "	12	1	1	1	1	S	Y	804	804	670		
	Spanish Gr 7 Math Answer Book	" "	12	1	1	1	1	S	Y	805	805	671		
	Spanish Gr 8 Math Answer Book	" "	12	1	1	1	1	S	Y	796	796	663		
	Spanish Algebra I Answer Book	" "	16	1	1	1	1	S	Y	197	197	164		
	Spanish Algebra II Answer Book	" "	16	1	1	1	1	S	Y	186	186	155		
	Spanish Geometry Answer Book	" "	16	1	1	1	1	S	Y	170	170	142		
	Spanish Int I Answer Book	" "	16	1	1	1	1	S	Y	65	65	54		
	Spanish Int II Answer Book	" "	16	1	1	1	1	S	Y	62	62	52		
	Spanish Int III Answer Book	" "	16	1	1	1	1	S	Y	56	56	47		
Scannable Headers			2	1	1	1	1			90,460	82,236	74,760		
	School/Classroom Header	scannable sheet, 166 (6x6 layout), 9.5x11, mark reflex, black plus one additional color, 60# offset, continuous, lithocode front/back	2	1	1	1	1	S	Y	90,460	82,236	74,760	Custom	Placed in District and/or School Coordinator kit
Non-scannable Test Books/Scripts			1,184	341	321	194	179	-	-	5,837,750	5,837,750	-		
ELA--PBA	Grade 4 ELA -- Test Booklet	8-3/8X10-7/8 booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color + black on cover, one color + black internal	48	6	6	5	4	NS	Y	498,538	498,538		Custom to order	Custom to secure order assignment
	Grade 5 ELA -- Test Booklet	" "	48	6	6	6	4	NS	Y	496,505	496,505		"	
	Grade 6 ELA -- Test Booklet	" "	48	6	6	6	4	NS	Y	492,246	492,246		"	
	Grade 7 ELA --Test Booklet	" "	48	6	6	6	4	NS	Y	492,937	492,937		"	
	Grade 8 ELA -- Test Booklet	" "	48	6	6	6	4	NS	Y	486,968	486,968		"	
	Grade 9 ELA --Test Booklet	" "	48	4	4	4	2	NS	Y	160,614	160,614		"	
	Grade 10 ELA --Test Booklet	" "	48	4	4	4	2	NS	Y	151,709	151,709		"	
	Grade 11 ELA -- Test Booklet	" "	48	4	4	4	2	NS	Y	139,059	139,059		"	
	Grade 4 Math Test Booklet	" "	32	25	25	15	15	NS	Y	497,826	497,826		"	
	Grade 5 Math Test Booklet	" "	32	25	25	15	15	NS	Y	495,795	495,795		"	

	Document Title	Document Description	# Pgs	# of Fms	Y2	Y3	Y4	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
MATH--PBA	Grade 6 Math Test Booklet	" "	32	25	25	15	15	NS	Y	491,543	491,543		"	
	Grade 7 Math Test Booklet	" "	32	25	25	15	15	NS	Y	492,233	492,233		"	
	Grade 8 Math Test Booklet	" "	32	25	25	15	15	NS	Y	486,272	486,272		"	
	Algebra I Test Booklet	" "	40	25	25	14	14	NS	Y	120,284	120,284		"	
	Algebra II Test Booklet	" "	40	33	25	18	18	NS	Y	113,611	113,611		"	
	Geometry Test Booklet	" "	40	25	25	14	14	NS	Y	104,155	104,155		"	
	Int I Test Booklet	" "	40	24	21	7	7	NS	Y	40,094	40,094		"	
	Int II Test Booklet	" "	40	24	21	7	7	NS	Y	37,876	37,876		"	
	Int III Test Booklet	" "	40	32	26	7	7	NS	Y	34,718	34,718		"	
	Spanish Gr 4 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	814	814			
	Spanish Gr 5 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	811	811			
	Spanish Gr 6 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	804	804			
	Spanish Gr 7 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	805	805			
	Spanish Gr 8 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	796	796			
	Spanish Algebra I Test Booklet	" "	40	1	1	1	1	NS	Y	197	197			
	Spanish Algebra II Test Booklet	" "	40	1	1	1	1	NS	Y	186	186			
	Spanish Geometry Test Booklet	" "	40	1	1	1	1	NS	Y	170	170			
	Spanish Int I Test Booklet	" "	40	1	1	1	1	NS	Y	65	65			
	Spanish Int II Test Booklet	" "	40	1	1	1	1	NS	Y	62	62			
	Spanish Int III Test Booklet	" "	40	1	1	1	1	NS	Y	56	56			
Large Print			1,320	33	33	33	33	-	-	25,313	25,313	-		
	Large Print Grade 3 ELA Test Book	large print book, 14X17, 150% photo enlargement, black on white 50# stock, coil bind (minimum 18-pt type)	56	1	1	1	1	NS	Y	1,827	1,827		1's, Kit	Kits to include instruction memo, corresponding form 1 test book & matching scan A/doc
	Large Print Grade 4 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,845	1,845		"	" "
	Large Print Grade 5 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,838	1,838		"	" "
	Large Print Grade 6 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,822	1,822		"	" "
	Large Print Grade 7 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,824	1,824		"	" "
	Large Print Grade 8 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,802	1,802		"	" "
	Large Print Grade 9 ELA Test Book	" "	48	1	1	1	1	NS	Y	593	593		1's, Kit	Kits to include instruction memo, corresponding form 1 test book & matching scan A/doc
	Large Print Grade 10 ELA Test Book	" "	48	1	1	1	1	NS	Y	562	562		"	" "
	Large Print Grade 11 ELA Test Book	" "	48	1	1	1	1	NS	Y	516	516		"	" "
	Large Print Grade 3 Math Test Book	" "	40	1	1	1	1	NS	Y	1,827	1,827		"	" "
	Large Print Grade 4 Math Test Book	" "	32	1	1	1	1	NS	Y	1,845	1,845		"	" "
	Large Print Grade 5 Math Test Book	" "	32	1	1	1	1	NS	Y	1,838	1,838		"	" "
	Large Print Grade 6 Math Test Book	" "	32	1	1	1	1	NS	Y	1,822	1,822		"	" "
	Large Print Grade 7 Math Test Book	" "	32	1	1	1	1	NS	Y	1,824	1,824		"	" "
	Large Print Grade 8 Math Test Book	" "	32	1	1	1	1	NS	Y	1,802	1,802		"	" "
	Large Print Algebra I Test Book	" "	40	1	1	1	1	NS	Y	446	446		"	" "
	Large Print Algebra II Test Book	" "	40	1	1	1	1	NS	Y	420	420		"	" "
	Large Print Geometry Test Book	" "	40	1	1	1	1	NS	Y	385	385		"	" "
	Large Print Int I Test Book	" "	40	1	1	1	1	NS	Y	149	149		"	" "
	Large Print Int II Test Book	" "	40	1	1	1	1	NS	Y	140	140		"	" "
	Large Print Int III Test Book	" "	40	1	1	1	1	NS	Y	128	128		"	" "
	Lg Print Spanish Gr 3 Mathematics Test Booklet	" "	40	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Gr 4 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Gr 5 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Gr 6 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			

	Document Title	Document Description	# Pgs	# of Fms	Y2	Y3	Y4	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
	Lg Print Spanish Gr 7 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Gr 8 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Algebra I Test Booklet	" "	40	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Algebra II Test Booklet	" "	40	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Geometry Test Booklet	" "	40	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Int I Test Booklet	" "	40	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Int II Test Booklet	" "	40	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Int III Test Booklet	" "	40	1	1	1	1	NS	Y	5	5			
Braille			2,200	21	21	21	21	-	-	2,405	2,405	-		
	Braille Grade 3 ELA Test Book	Contracted Braille: 11-1/2x11" Manila 100# stock	140	1	1	1	1	NS	Y	174	174		1's, Kit	Kits to include instructions and matching scan A/doc
	Braille Grade 4 ELA Test Book	" "	120	1	1	1	1	NS	Y	176	176		"	" "
	Braille Grade 5 ELA Test Book	" "	120	1	1	1	1	NS	Y	175	175		"	" "
	Braille Grade 6 ELA Test Book	" "	120	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 7 ELA Test Book	" "	120	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 8 ELA Test Book	" "	120	1	1	1	1	NS	Y	172	172		"	" "
	Braille Grade 9 ELA Test Book	" "	120	1	1	1	1	NS	Y	56	56		1's, Kit	Kits to include instructions and matching scan A/doc
	Braille Grade 10 ELA Test Book	" "	120	1	1	1	1	NS	Y	53	53		"	" "
	Braille Grade 11 ELA Test Book	" "	120	1	1	1	1	NS	Y	50	50		"	" "
	Braille Grade 3 Math Test Book	" "	100	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 4 Math Test Book	" "	80	1	1	1	1	NS	Y	176	176		"	" "
	Braille Grade 5 Math Test Book	" "	80	1	1	1	1	NS	Y	175	175		"	" "
	Braille Grade 6 Math Test Book	" "	80	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 7 Math Test Book	" "	80	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 8 Math Test Book	" "	80	1	1	1	1	NS	Y	172	172		"	" "
	Braille Algebra I Test Book	" "	100	1	1	1	1	NS	Y	43	43		"	" "
	Braille Print Algebra II Test Book	" "	100	1	1	1	1	NS	Y	39	39		"	" "
	Braille Print Geometry Test Book	" "	100	1	1	1	1	NS	Y	36	36		"	" "
	Braille Print Int I Test Book	" "	100	1	1	1	1	NS	Y	14	14		"	" "
	Braille Print Int II Test Book	" "	100	1	1	1	1	NS	Y	13	13		"	" "
	Braille Print Int III Test Book	" "	100	1	1	1	1	NS	Y	11	11		"	" "
Other Non-Scannable Printed Materials (Manuals, etc.)			1,504	10	10	10	10	-	-	356,143	323,766	-		
	Test Coordinator Manual (Paper)	8-1/2 x 11 self cover booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	96	1	1	1	1	NS	Y	16,183	14,712		1's	Placed in District and School Coord Kits
	Test Coordinator Manual (CBA)	" "	96	1	1	1	1	NS	Y	16,183	14,712		1's	Placed in District and School Coord Kits
	Test Admin Manual (Paper) Gr 3-5 ELA & Mathematics	8-1/2 x 11 self cover booklet, perfect-bound, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	184	1	1	1	1	NS	Y	62,395	56,723		TBD - 5's & 1's or other	
	Test Admin Manual (Paper) Gr 6-8 ELA & Mathematics	" "	184	1	1	1	1	NS	Y	61,690	56,082		"	
	Test Admin Manual (Paper) - Gr 9-11 ELA	8-1/2 x 11 self cover booklet, perfect-bound, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	144	1	1	1	1	NS	Y	18,915	17,196		"	
	Test Admin Manual (Paper) Gr 9-11 Mathematics	" "	144	1	1	1	1	NS	Y	18,888	17,171		"	
	Test Admin Manual (CBA) Gr 3-5 ELA & Mathematics	" "	184	1	1	1	1	NS	Y	62,395	56,723		"	
	Test Admin Manual (CBA) Gr 6-8 ELA & Mathematics	" "	184	1	1	1	1	NS	Y	61,690	56,082		"	
	Test Admin Manual (CBA) - Gr 9-11 ELA	" "	144	1	1	1	1	NS	Y	18,915	17,196		"	
	Test Admin Manual (CBA) - Gr 9-11 Mathematics	" "	144	1	1	1	1	NS	Y	18,888	17,171		"	

	Document Title	Document Description	# Pgs	# of Fms	Y2	Y3	Y4	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
Printed & Other Non-Stock Materials (CD. Cassettes, Etc.)			143	83	80	80	80	-	-	8,055,895	8,055,895	-		
	Grade 3 Math Reference Sheet	8.5 x 11 Single sheet, printed black simplex on 50# white offset, Wrap in 20's	2	1	1	1	1	NS	Y	517,878	517,878		20's or TBD	Will come from the vendor wrapped in 20's or other TBD size
	Grade 4 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	522,717	522,717		"	" "
	Grade 5 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	520,585	520,585		"	" "
	Grade 6 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	516,120	516,120		"	" "
	Grade 7 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	516,844	516,844		"	" "
	Grade 8 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	510,585	510,585		"	" "
	Algebra I Reference sheet	" "	2	1	1	1	1	NS	Y	126,298	126,298		"	" "
	Algebra II Reference sheet	" "	2	1	1	1	1	NS	Y	119,292	119,292		"	" "
	Geometry Reference sheet	" "	2	1	1	1	1	NS	Y	109,362	109,362		"	" "
	Int I Reference sheet	" "	2	1	1	1	1	NS	Y	42,099	42,099		"	" "
	Int II Reference sheet	" "	2	1	1	1	1	NS	Y	39,769	39,769		"	" "
	Int III Reference sheet	" "	2	1	1	1	1	NS	Y	36,454	36,454		"	" "
	Spanish Gr 3 Math Reference Sheet	8.5 x 11 Single sheet, printed black simplex on 50# white offset, Wrap in 20's	2	1	1	1	1	NS	Y	847	847			
	Spanish Gr 4 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	854	854			
	Spanish Gr 5 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	852	852			
	Spanish Gr 6 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	844	844			
	Spanish Gr 7 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	845	845			
	Spanish Gr 8 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	835	835			
	Spanish Algebra I Reference sheet	" "	2	1	1	1	1	NS	Y	207	207			
	Spanish Algebra II Reference sheet	" "	2	1	1	1	1	NS	Y	195	195			
	Spanish Geometry Reference sheet	" "	2	1	1	1	1	NS	Y	179	179			
	Spanish Int I Reference sheet	" "	2	1	1	1	1	NS	Y	68	68			
	Spanish Int II Reference sheet	" "	2	1	1	1	1	NS	Y	66	66			
	Spanish Int III Reference sheet	" "	2	1	1	1	1	NS	Y	59	59			
LP	Lg Print - Grade 3 Math Reference Sheet	Single Lg Print 14X17 sheet, 150% photo enlargement, black on white 50# stock (minimum 18-pt type)	2	1	1	1	1	NS		1,827	1,827		1's, Kit	Part of Lg Print kit
LP	Lg Print - Grade 4 Math Reference Sheet	" "	2	1	1	1	1	NS		1,845	1,845		"	" "
LP	Lg Print - Grade 5 Math Reference Sheet	" "	2	1	1	1	1	NS		1,838	1,838		"	" "
LP	Lg Print - Grade 6 Math Reference Sheet	" "	2	1	1	1	1	NS		1,822	1,822		"	" "
LP	Lg Print - Grade 7 Math Reference Sheet	" "	2	1	1	1	1	NS		1,824	1,824		"	" "
LP	Lg Print - Grade 8 Math Reference Sheet	" "	2	1	1	1	1	NS		1,802	1,802		"	" "
LP	Lg Print - Algebra I Reference sheet	Single Lg Print 14X17 sheet, 150% photo enlargement, black on white 50# stock (minimum 18-pt type)	2	1	1	1	1	NS		446	446		1's, Kit	Part of Lg Print kit
LP	Lg Print - Algebra II Reference sheet	" "	2	1	1	1	1	NS		420	420		"	" "
LP	Lg Print - Geometry Reference sheet	" "	2	1	1	1	1	NS		385	385		"	" "
LP	Lg Print - Int I Reference sheet	" "	2	1	1	1	1	NS		149	149		"	" "
LP	Lg Print - Int II Reference sheet	" "	2	1	1	1	1	NS		140	140		"	" "
LP	Lg Print - Int III Reference sheet	" "	2	1	1	1	1	NS		128	128		"	" "
LP	Spanish Lg Print - Grade 3 Math Reference Sheet	Single Lg Print 14X17 sheet, 150% photo enlargement, black on white 50# stock (minimum 18-pt type)	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Grade 4 Math Reference Sheet	" "	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Grade 5 Math Reference Sheet	" "	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Grade 6 Math Reference Sheet	" "	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Grade 7 Math Reference Sheet	" "	2	1	1	1	1	NS		2	2			

	Document Title	Document Description	# Pgs	# of Fms	Y2	Y3	Y4	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
LP	Spanish Lg Print - Grade 8 Math Reference Sheet	" "	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Algebra I Reference sheet	Single Lg Print 14X17 sheet, 150% photo enlargement, black on white 50# stock (minimum 18-pt type)	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Algebra II Reference sheet	" "	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Geometry Reference sheet	" "	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Int I Reference sheet	" "	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Int II Reference sheet	" "	2	1	1	1	1	NS		2	2			
LP	Spanish Lg Print - Int III Reference sheet	" "	2	1	1	1	1	NS		2	2			
BR	BRAILLE - Grade 3 Math Reference Sheet	Contracted Braille: 11-1/2x11" sheet, Manila 100# stock	2	1	1	1	1	NS	Y	174	174		1's, Kit	Part of Braille kit
BR	BRAILLE - Grade 4 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	176	176		"	" "
BR	BRAILLE - Grade 5 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	175	175		"	" "
BR	BRAILLE - Grade 6 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	174	174		"	" "
BR	BRAILLE - Grade 7 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	174	174		"	" "
BR	BRAILLE - Grade 8 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	172	172		"	" "
BR	BRAILLE - Algebra I Reference sheet	Contracted Braille: 11-1/2x11" sheet, Manila 100# stock	2	1	1	1	1	NS	Y	43	43		1's, Kit	Part of Braille kit
BR	BRAILLE - Algebra II Reference sheet	" "	2	1	1	1	1	NS	Y	39	39		"	" "
BR	BRAILLE - Geometry Reference sheet	" "	2	1	1	1	1	NS	Y	36	36		"	" "
BR	BRAILLE - Int I Reference sheet	" "	2	1	1	1	1	NS	Y	14	14		"	" "
BR	BRAILLE - Int II Reference sheet	" "	2	1	1	1	1	NS	Y	13	13		"	" "
BR	BRAILLE - Int III Reference sheet	" "	2	1	1	1	1	NS	Y	11	11		"	" "
	Ruler -Grade 3	Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	0.2	1	1	1	1	NS	N	493,218	493,218		TBD	
	Ruler - Grade 4-5	Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	993,621	993,621		TBD	
	Ruler - Grade - 6-HS	Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	1,920,785	1,920,785		TBD	
	Protractor - Grade 4-5 only	Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1	1	1	1	NS	N	993,621	993,621		TBD	
LP	LG Print - Ruler -Grade 3	LP version to support: Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	1,827	1,827		1's, Kit	Part of Lg Print kit
LP	LG Print - Ruler - Grade 4-5	LP version to support: Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	1	1	1	NS	N	3,682	3,682		"	" "
LP	LG Print - Ruler - Grade - 6-HS	LP version to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1				NS	N	7,116	7,116		1's, Kit	Part of Lg Print kit
LP	LG Print - Protractor - Grade 4-5 only	LP version to support: Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1				NS	N	3,682	3,682		"	" "

	Document Title	Document Description	# Pgs	# of Fms	Y2	Y3	Y4	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
	Spanish Ruler -Grade 3	Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	0.2	1	1	1	1	NS	Y	847	847			
	Spanish Ruler - Grade 4-5	Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	Y	1,706	1,706			
	Spanish Ruler - Grade - 6-HS	Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	Y	7,116	7,116			
LP	Spanish - Protractor - Grade 4-5 only	Spanish LP version to support: Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1	1	1	1	NS	Y	1,706	1,706			
LP	Spanish LG Print - Ruler - Grade - 3	LP version to support: Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	0.4	1	1	1	1			2	2			
LP	Spanish LG Print - Ruler - Grade - 4-5	LP version to support: Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	1	1	1			4	4			
LP	Spanish Ruler LG Print - Ruler - Grade - 6-HS	Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	1	1	1			18	18			
LP	Spanish LG Print - Protractor - Grade 4-5 only	Spanish LP version to support: Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1	1	1	1			4	4			
BR	BRAILLE - Ruler -Grade 3	Brailled to support: Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock.	0.2	1	1	1	1	NS	N	174	174		1's, Kit	Part of Braille kit
BR	BRAILLE - Ruler - Grade 4-5	Brailled to support: Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	351	351		"	" "
BR	BRAILLE - Ruler - Grade - 6-HS	Brailled to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	676	676		1's, Kit	Part of Braille kit
BR	BRAILLE - Protractor - Grade 4-5 only	Brailled to support: Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1	1	1	1	NS	N	351	351		"	" "
	Large Print Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	1	1	1	NS	Y	25,253	25,253		"	Insert in Lg Print kit assemby
	Spanish Lg Print instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	1	1	1	NS	Y	24	24			
	Braille Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1				NS	N	2,405	2,405		"	Insert in Braille kit assemby
	Read Aloud Kit	Provide uncollated (test book) units for use with students with special accommodations to have the test read aloud in small groups.									-			Nothing known related to any additonal materials.
Miscellaneous Materials (Stock)			3	2	-	-	-	-	-	7,845,442	8,410,141	-		
	Pre-ID Labels	LABEL, PRE-ID, 3900, 3.5X1 7/16 TONER BACK	0.2	1				NS	Y	6,825,697	6,825,697			
	Paper Bands	standard 3"x27" brown kraft self-sealing bands	1	1				NS	N	455,046	455,046			
	Stock Colored Return labels	3x5 colored return labels, 2-up, laser compatible	0.5					NS	N	188,233	188,233			
	Stock Outbound Shipping labels	3x5 white shipping labels, 2-up, laser compatible	0.5					NS	N	188,233	188,233			
	UPS Labels	As needed for carrier. Stock	0.5					NS	N	188,233	188,233			

Y1-Y4 Traditional Spr 3-11 EOY

12,460 Schools
5,647 Distribution Points
2,252 Systems/Districts

	Document Title	Document Description	# Pgs	# of Fms	Y2	Y3	Y4	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
Scannable Materials			344	262	262	138	138	-	-	6,825,697	6,825,697	6,500,000		
ELA--EOY	Grade 3 ELA -- Consumable Test Booklet	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	48	43	43	29	29	S	Y	493,923	493,923	470,403	Custom to order need	Custom to secure order assignment
	Grade 4 ELA -- Answer Doc	" "	2	1	1	1	1	S	Y	498,538	498,538	474,798	20's, 5's - or other TBD	
	Grade 5 ELA -- Answer Doc	" "	2	1	1	1	1	S	Y	496,505	496,505	472,862	"	
	Grade 6 ELA -- Answer Doc	" "	2	1	1	1	1	S	Y	492,246	492,246	468,806	"	
	Grade 7 ELA -- Answer Doc	" "	2	1	1	1	1	S	Y	492,937	492,937	469,464	"	
	Grade 8 ELA -- Answer Doc	" "	2	1	1	1	1	S	Y	486,968	486,968	463,779	"	
	Grade 9 ELA -- Answer Doc	" "	2	1	1	1	1	S	Y	160,614	160,614	152,966	"	
	Grade 10 ELA -- Answer Doc	" "	2	1	1	1	1	S	Y	151,709	151,709	144,485	"	
	Grade 11 ELA -- Answer Doc	" "	2	1	1	1	1	S	Y	139,059	139,059	132,437	"	
MATH--EOY	Grade 3 Math Consumable Test Booklet	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	28	39	39	20	20	S	Y	493,218	493,218	469,731	Custom to order	Custom to secure order assignment
	Grade 4 Math Answer Book	" "	8	14	14	7	7	S	Y	497,826	497,826	474,120	20's, 5's - or other TBD	
	Grade 5 Math Answer Book	" "	8	14	14	7	7	S	Y	495,795	495,795	472,186	"	
	Grade 6 Math Answer Book	" "	8	14	14	7	7	S	Y	491,543	491,543	468,136	"	
	Grade 7 Math Answer Book	" "	8	13	13	7	7	S	Y	492,233	492,233	468,793	"	
	Grade 8 Math Answer Book	" "	8	14	14	7	7	S	Y	486,272	486,272	463,116	"	
	Algebra I Answer Book	" "	12	16	16	8	8	S	Y	120,284	120,284	114,556	"	
	Algebra II Answer Book	" "	12	15	15	7	7	S	Y	113,611	113,611	108,201	"	
	Geometry Answer Book	" "	12	15	15	7	7	S	Y	104,155	104,155	99,195	"	
	Int I Answer Book	" "	12	15	15	4	4	S	Y	40,094	40,094	38,185	"	
	Int II Answer Book	" "	12	15	15	4	4	S	Y	37,876	37,876	36,072	"	
	Int III Answer Book	" "	12	15	15	4	4	S	Y	34,718	34,718	33,065	"	
	Spanish Gr 3 Math Consumable Test Booklet	" "	28	1	1	1	1	S	Y	806	806	672	"	
	Spanish Gr 4 Math Answer Book	" "	8	1	1	1	1	S	Y	814	814	678	"	
	Spanish Gr 5 Math Answer Book	" "	8	1	1	1	1	S	Y	811	811	676	"	
	Spanish Gr 6 Math Answer Book	" "	8	1	1	1	1	S	Y	804	804	670	"	
	Spanish Gr 7 Math Answer Book	" "	8	1	1	1	1	S	Y	805	805	671	"	
	Spanish Gr 8 Math Answer Book	" "	8	1	1	1	1	S	Y	796	796	663	"	
	Spanish Algebra I Answer Book	" "	12	1	1	1	1	S	Y	197	197	164	"	
	Spanish Algebra II Answer Book	" "	12	1	1	1	1	S	Y	186	186	155	"	
	Spanish Geometry Answer Book	" "	12	1	1	1	1	S	Y	170	170	142	"	
	Spanish Int I Answer Book	" "	12	1	1	1	1	S	Y	65	65	54	"	
	Spanish Int II Answer Book	" "	12	1	1	1	1	S	Y	62	62	52	"	
	Spanish Int III Answer Book	" "	12	1	1	1	1	S	Y	56	56	47	"	
Scannable Headers			2	1	1	1	1	-	-	90,460	82,236	74,760		
	School/Classroom Header	scannable sheet, 166 (6x6 layout), 9.5x11, mark reflex, black plus one additional color, 60# offset, continuous, lithocode front/back	2	1	1	1	1	S	Y	90,460	82,236	74,760	Custom	Placed in District and/or School Coordinator kit
Non-scannable Test Books/Scripts			1,056	862	862	484	484	-	-	5,837,750	5,837,750	-		
ELA--EOY	Grade 4 ELA -- Test Booklet	8-3/8X10-7/8 booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color + black on cover, one color + black internal	48	46	46	31	31	NS	Y	498,538	498,538		Custom to order	Custom to secure order assignment
	Grade 5 ELA -- Test Booklet	" "	48	46	46	31	31	NS	Y	496,505	496,505		"	
	Grade 6 ELA -- Test Booklet	" "	48	51	51	34	34	NS	Y	492,246	492,246		"	
	Grade 7 ELA --Test Booklet	" "	48	51	51	34	34	NS	Y	492,937	492,937		"	
	Grade 8 ELA -- Test Booklet	" "	48	51	51	34	34	NS	Y	486,968	486,968		"	
	Grade 9 ELA --Test Booklet	" "	48	47	47	38	38	NS	Y	160,614	160,614		"	
	Grade 10 ELA --Test Booklet	" "	48	47	47	38	38	NS	Y	151,709	151,709		"	
	Grade 11 ELA -- Test Booklet	" "	48	45	45	34	34	NS	Y	139,059	139,059		"	

MATH--EOY	Grade 4 Math Test Booklet	" "	24	40	40	20	20	NS	Y	497,826	497,826		"	
	Grade 5 Math Test Booklet	" "	24	40	40	21	21	NS	Y	495,795	495,795		"	
	Grade 6 Math Test Booklet	" "	32	41	41	21	21	NS	Y	491,543	491,543		"	
	Grade 7 Math Test Booklet	" "	32	39	39	20	20	NS	Y	492,233	492,233		"	
	Grade 8 Math Test Booklet	" "	32	40	40	20	20	NS	Y	486,272	486,272		"	
	Algebra I Test Booklet	" "	32	47	47	24	24	NS	Y	120,284	120,284		"	
	Algebra II Test Booklet	" "	32	44	44	21	21	NS	Y	113,611	113,611		"	
	Geometry Test Booklet	" "	32	44	44	21	21	NS	Y	104,155	104,155		"	
	Int I Test Booklet	" "	32	44	44	10	10	NS	Y	40,094	40,094		"	
	Int II Test Booklet	" "	32	44	44	10	10	NS	Y	37,876	37,876		"	
	Int III Test Booklet	" "	32	44	44	11	11	NS	Y	34,718	34,718		"	
	Spanish Gr 4 Mathematics Test Booklet	" "	24	1	1	1	1	NS	Y	814	814			
	Spanish Gr 5 Mathematics Test Booklet	" "	24	1	1	1	1	NS	Y	811	811			
	Spanish Gr 6 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	804	804			
	Spanish Gr 7 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	805	805			
	Spanish Gr 8 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	796	796			
	Spanish Algebra I Test Booklet	" "	32	1	1	1	1	NS	Y	197	197			
	Spanish Algebra II Test Booklet	" "	32	1	1	1	1	NS	Y	186	186			
	Spanish Geometry Test Booklet	" "	32	1	1	1	1	NS	Y	170	170			
	Spanish Int I Test Booklet	" "	32	1	1	1	1	NS	Y	65	65			
	Spanish Int II Test Booklet	" "	32	1	1	1	1	NS	Y	62	62			
	Spanish Int III Test Booklet	" "	32	1	1	1	1	NS	Y	56	56			
Large Print			1,160	33	33	33	33	-	-	25,313	25,313	-		
	Large Print Grade 3 ELA Test Book	large print book, 14X17, 150% photo enlargement, black on white 50# stock, coil bind (minimum 18-pt type)	48	1	1	1	1	NS	Y	1,827	1,827		1's, Kit	Kits to include instruction memo, corresponding form 1 test book & matching scan A/doc
	Large Print Grade 4 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,845	1,845		"	" "
	Large Print Grade 5 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,838	1,838		"	" "
	Large Print Grade 6 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,822	1,822		"	" "
	Large Print Grade 7 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,824	1,824		"	" "
	Large Print Grade 8 ELA Test Book	" "	48	1	1	1	1	NS	Y	1,802	1,802		"	" "
	Large Print Grade 9 ELA Test Book	" "	48	1	1	1	1	NS	Y	593	593		1's, Kit	Kits to include instruction memo, corresponding form 1 test book & matching scan A/doc
	Large Print Grade 10 ELA Test Book	" "	48	1	1	1	1	NS	Y	562	562		"	" "
	Large Print Grade 11 ELA Test Book	" "	48	1	1	1	1	NS	Y	516	516		"	" "
	Large Print Grade 3 Math Test Book	" "	28	1	1	1	1	NS	Y	1,827	1,827		"	" "
	Large Print Grade 4 Math Test Book	" "	24	1	1	1	1	NS	Y	1,845	1,845		"	" "
	Large Print Grade 5 Math Test Book	" "	24	1	1	1	1	NS	Y	1,838	1,838		"	" "
	Large Print Grade 6 Math Test Book	" "	32	1	1	1	1	NS	Y	1,822	1,822		"	" "
	Large Print Grade 7 Math Test Book	" "	32	1	1	1	1	NS	Y	1,824	1,824		"	" "
	Large Print Grade 8 Math Test Book	" "	32	1	1	1	1	NS	Y	1,802	1,802		"	" "
	Large Print Algebra I Test Book	" "	32	1	1	1	1	NS	Y	446	446		"	" "
	Large Print Algebra II Test Book	" "	32	1	1	1	1	NS	Y	420	420		"	" "
	Large Print Geometry Test Book	" "	32	1	1	1	1	NS	Y	385	385		"	" "
	Large Print Int I Test Book	" "	32	1	1	1	1	NS	Y	149	149		"	" "
	Large Print Int II Test Book	" "	32	1	1	1	1	NS	Y	140	140		"	" "
	Large Print Int III Test Book	" "	32	1	1	1	1	NS	Y	128	128		"	" "
	Lg Print Spanish Gr 3 Mathematics Test Booklet	" "	28	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Gr 4 Mathematics Test Booklet	" "	24	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Gr 5 Mathematics Test Booklet	" "	24	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Gr 6 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			

	Lg Print Spanish Gr 7 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Gr 8 Mathematics Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Algebra I Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Algebra II Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Geometry Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Int I Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Int II Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
	Lg Print Spanish Int III Test Booklet	" "	32	1	1	1	1	NS	Y	5	5			
Braille			1,990	21	21	21	21	-	-	2,405	2,405	-		
	Braille Grade 3 ELA Test Book	Contracted Braille: 11-1/2x11" Manila 100# stock	120	1	1	1	1	NS	Y	174	174		1's, Kit	Kits to include instructions and matching scan A/doc
	Braille Grade 4 ELA Test Book	" "	120	1	1	1	1	NS	Y	176	176		"	" "
	Braille Grade 5 ELA Test Book	" "	120	1	1	1	1	NS	Y	175	175		"	" "
	Braille Grade 6 ELA Test Book	" "	120	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 7 ELA Test Book	" "	120	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 8 ELA Test Book	" "	120	1	1	1	1	NS	Y	172	172		"	" "
	Braille Grade 9 ELA Test Book	" "	120	1	1	1	1	NS	Y	56	56		1's, Kit	Kits to include instructions and matching scan A/doc
	Braille Grade 10 ELA Test Book	" "	120	1	1	1	1	NS	Y	53	53		"	" "
	Braille Grade 11 ELA Test Book	" "	120	1	1	1	1	NS	Y	50	50		"	" "
	Braille Grade 3 Math Test Book	" "	70	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 4 Math Test Book	" "	60	1	1	1	1	NS	Y	176	176		"	" "
	Braille Grade 5 Math Test Book	" "	60	1	1	1	1	NS	Y	175	175		"	" "
	Braille Grade 6 Math Test Book	" "	80	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 7 Math Test Book	" "	80	1	1	1	1	NS	Y	174	174		"	" "
	Braille Grade 8 Math Test Book	" "	80	1	1	1	1	NS	Y	172	172		"	" "
	Braille Algebra I Test Book	" "	80	1	1	1	1	NS	Y	43	43		"	" "
	Braille Print Algebra II Test Book	" "	80	1	1	1	1	NS	Y	39	39		"	" "
	Braille Print Geometry Test Book	" "	80	1	1	1	1	NS	Y	36	36		"	" "
	Braille Print Int I Test Book	" "	80	1	1	1	1	NS	Y	14	14		"	" "
	Braille Print Int II Test Book	" "	80	1	1	1	1	NS	Y	13	13		"	" "
	Braille Print Int III Test Book	" "	80	1	1	1	1	NS	Y	11	11		"	" "
Other Non-Scannable Printed Materials (Manuals, etc.)			-	-	-	-	-	-	-	-	-	-		
Printed & Other Non-Stock Materials (CD. Cassettes, Etc.)			141	82	81	81	81	-	-	9,321,174	9,321,174	-		
	Grade 3 Math Reference Sheet	8.5 x 11 Single sheet, printed black simplex on 50# white offset, Wrap in 20's	2	1	1	1	1	NS	Y	517,878	517,878			
	Grade 4 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	523,465	523,465			
	Grade 5 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	521,330	521,330			
	Grade 6 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	516,859	516,859			
	Grade 7 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	517,584	517,584			
	Grade 8 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	511,316	511,316			
	Algebra I Reference sheet	" "	2	1	1	1	1	NS	Y	168,645	168,645			
	Algebra II Reference sheet	" "	2	1	1	1	1	NS	Y	159,295	159,295			
	Geometry Reference sheet	" "	2	1	1	1	1	NS	Y	146,012	146,012		"	" "
	Int II Reference sheet	" "	2	1	1	1	1	NS	Y	522,717	522,717		"	" "
	Int III Reference sheet	" "	2	1	1	1	1	NS	Y	520,585	520,585		"	" "
LP	Lg Print - Grade 3 Math Reference Sheet	Single Lg Print 14X17 sheet, 150% photo enlargement, black on white 50# stock (minimum 18-pt type)	2	1						1,918	1,918		1's, Kit	Part of Lg Print kit
LP	Lg Print - Grade 4 Math Reference Sheet	" "	2	1	1	1	1	NS		1,937	1,937		"	" "
LP	Lg Print - Grade 5 Math Reference Sheet	" "	2	1	1	1	1	NS		1,929	1,929		"	" "
LP	Lg Print - Grade 6 Math Reference Sheet	" "	2	1	1	1	1	NS		1,913	1,913		"	" "
LP	Lg Print - Grade 7 Math Reference Sheet	" "	2	1	1	1	1	NS		1,915	1,915		"	" "
LP	Lg Print - Grade 8 Math Reference Sheet	" "	2	1	1	1	1	NS		1,892	1,892		"	" "
LP	Lg Print - Algebra I Reference sheet	" "	2	1	1	1	1	NS		469	469		"	" "
LP	Lg Print - Algebra II Reference sheet	" "	2	1	1	1	1	NS		441	441		"	" "
LP	Lg Print - Geometry Reference sheet	" "	2	1	1	1	1	NS		405	405		"	" "
LP	Lg Print - Int I Reference sheet	" "	2	1	1	1	1	NS		157	157		"	" "
LP	Lg Print - Int II Reference sheet	" "	2	1	1	1	1	NS		147	147		"	" "
LP	Lg Print - Int III Reference sheet	" "	2	1	1	1	1	NS		135	135		"	" "

	Spanish Gr 3 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	847	847			
	Spanish Gr 4 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	854	854			
	Spanish Gr 5 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	852	852			
	Spanish Gr 6 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	844	844			
	Spanish Gr 7 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	845	845			
	Spanish Gr 8 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	835	835			
	Spanish Algebra I Reference sheet	" "	2	1	1	1	1	NS	Y	207	207			
	Spanish Algebra II Reference sheet	" "	2	1	1	1	1	NS	Y	195	195			
	Spanish Geometry Reference sheet	" "	2	1	1	1	1	NS	Y	179	179			
	Spanish Int I Reference sheet	" "	2	1	1	1	1	NS	Y	68	68			
	Spanish Int II Reference sheet	" "	2	1	1	1	1	NS	Y	66	66			
	Spanish Int III Reference sheet	" "	2	1	1	1	1	NS	Y	59	59			
BR	Spanish Lg Print - Grade 3 Math Reference Sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Grade 4 Math Reference Sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Grade 5 Math Reference Sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Grade 6 Math Reference Sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Grade 7 Math Reference Sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Grade 8 Math Reference Sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Algebra I Reference sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Algebra II Reference sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Geometry Reference sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Int I Reference sheet	" "	2	1	1	1	1	NS		8	8			
BR	Spanish Lg Print - Int II Reference sheet	" "	2	1	1	1	1	NS		8	8			
	Spanish Lg Print - Int III Reference sheet	Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	2	1	1	1	1	NS		8	8			
BR	BRAILLE - Grade 3 Math Reference Sheet	Contracted Braille: 11-1/2x11" sheet, Manila 100# stock	2	1	1	1	1	NS	Y	174	174		1's, Kit	Part of Braille kit
BR	BRAILLE - Grade 4 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	176	176		"	" "
BR	BRAILLE - Grade 5 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	175	175		"	" "
BR	BRAILLE - Grade 6 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	174	174		"	" "
BR	BRAILLE - Grade 7 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	174	174		"	" "
BR	BRAILLE - Grade 8 Math Reference Sheet	" "	2	1	1	1	1	NS	Y	172	172		"	" "
BR	BRAILLE - Algebra I Reference sheet	" "	2	1	1	1	1	NS	Y	43	43		"	" "
BR	BRAILLE - Algebra II Reference sheet	" "	2	1	1	1	1	NS	Y	39	39		"	" "
BR	BRAILLE - Geometry Reference sheet	" "	2	1	1	1	1	NS	Y	36	36		"	" "
BR	BRAILLE - Int I Reference sheet	" "	2	1	1	1	1	NS	Y	14	14		"	" "
BR	BRAILLE - Int II Reference sheet	" "	2	1	1	1	1	NS	Y	13	13		"	" "
BR	BRAILLE - Int III Reference sheet	" "	2	1	1	1	1	NS	Y	11	11		"	" "
	Ruler -Grade 3	Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	0.2	1	1	1	1	NS	N	517,878	517,878		TBD	
	Ruler - Grade 4-5	Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	1,043,302	1,043,302		TBD	

	Ruler - Grade - 6-HS	Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	2,016,824	2,016,824		TBD	
	Protractor - Grade 4-5 only	Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1	1	1	1	NS	N	1,043,302	1,043,302		TBD	
LP	LG Print - Ruler -Grade 3	LP version to support: Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	1,918	1,918		1's, Kit	Part of Lg Print kit
LP	LG Print - Ruler - Grade 4-5	LP version to support: Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	1	1	1	NS	N	3,866	3,866		"	" "
LP	LG Print - Ruler - Grade - 6-HS	LP version to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	1	1	1	NS	N	7,116	7,116		"	" "
LP	LG Print - Protractor - Grade 4-5 only	LP version to support: Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1	1	1	1	NS	N	3,866	3,866		"	" "
	Spanish Ruler -Grade 3	Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	0.2	1	1	1	1	NS	N	847	847			
	Spanish Ruler - Grade 4-5	Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	1,706	1,706			
	Spanish Ruler - Grade - 6-HS	Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	3,299	3,299			
	Spanish Protractor - Grade 4-5 only	Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1	1	1	1	NS	N	1,706	1,706			
LP	Spanish LG Print - Ruler -Grade 3	LP version to support: Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	11	11			
LP	Spanish LG Print - Ruler - Grade 4-5	LP version to support: Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	1	1	1	NS	N	23	23			
LP	Spanish LG Print - Ruler - Grade - 6-HS	LP version to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	1	1	1	NS	N	101	101			
LP	Spanish LG Print - Protractor - Grade 4-5 only	LP version to support: Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1	1	1	1	NS	N	23	23			
BR	BRAILLE - Ruler -Grade 3	Brailled to support: Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock.	0.2	1	1	1	1	NS	N	174	174		1's, Kit	Part of Braille kit
BR	BRAILLE - Ruler - Grade 4-5	Brailled to support: Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	351	351		"	" "
BR	BRAILLE - Ruler - Grade - 6-HS	Brailled to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	1	1	1	NS	N	676	676		"	" "
BR	BRAILLE - Protractor - Grade 4-5 only	Brailled to support: Qty based on per protractor count - Clear velum or rigid clear plastic, one per student	2	1	1	1	1	NS	N	351	351		"	" "
	Large Print Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	1	1	1	NS	Y	25,253	25,253		"	Insert in Lg Print kit assemby
	Spanish Large Print Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	1	1	1	NS	N	90	90		"	Insert in Lg Print kit assemby
	Braille Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	1	1	1	NS	N	2,405	2,405		"	Insert in Braille kit assemby

	Read Aloud Kit	Provide uncollated (test book) units for use with students with special accommodations to have the test read aloud in small groups.									-			Nothing known related to any additonal materials.
Miscellaneous Materials (Stock)			3	-	-	-	-	-	-	7,693,206	8,033,148	-		
	Pre-ID Labels	LABEL, PRE-ID, 3900, 3.5X1 7/16 TONER BACK	0.2					NS	Y	6,825,697	6,825,697			
	Paper Bands	standard 3"x27" brown kraft self-sealing bands	1					NS	N	455,046	455,046			
	Stock Colored Return labels	3x5 colored return labels, 2-up, laser compatible	0.5					NS	N	137,488	137,488			
	Stock Outbound Shipping labels	3x5 white shipping labels, 2-up, laser compatible	0.5					NS	N	137,488	137,488			
	UPS Labels	As needed for carrier. Stock	0.5					NS	N	137,488	137,488			

Y1 Spr + Y2-4 Winter/Spr Block HS PBA's

1,145 Schools
948 Distribution Points
828 Systems/Districts

	Document Title	Document Description	# Pgs	# of Fms	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
Scannable Materials			228	18	-	-	262,670	262,670	250,000		
ELA--PBA	Grade 9 ELA -- Answer Book	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	12	1	S	Y	46,700	46,700	44,476	20's, 5's - or other TBD	
	Grade 10 ELA -- Answer Book	" "	12	1	S	Y	44,113	44,113	42,012	"	
	Grade 11 ELA -- Answer Book	" "	12	1	S	Y	40,437	40,437	38,512	"	
Math-PBA	Algebra I Answer Book	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	16	2	S	Y	34,977	34,977	33,311	"	
	Algebra II Answer Book	" "	16	2	S	Y	33,041	33,041	31,468	"	
	Geometry Answer Book	" "	16	2	S	Y	30,281	30,281	28,839	"	
	Int I Answer Book	" "	16	1	S	Y	11,659	11,659	11,104	"	
	Int II Answer Book	" "	16	1	S	Y	11,011	11,011	10,487	"	
	Int III Answer Book	" "	16	1	S	Y	10,094	10,094	9,613	"	
	Spanish Algebra I Answer Book	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	16	1			95	95	48		
	Spanish Algebra II Answer Book	" "	16	1			90	90	45		
	Spanish Geometry Answer Book	" "	16	1			82	82	41		
	Spanish Int I Answer Book	" "	16	1			32	32	16		
	Spanish Int II Answer Book	" "	16	1			30	30	15		
	Spanish Int III Answer Book	" "	16	1			28	28	14		
Scannable Headers			2	1			8,315	7,560	6,872		
	School/Classroom Header	scannable sheet, 166 (6x6 layout), 9.5x11, mark reflex, black plus one additional color, 60# offset, continuous, lithocode front/back	2	1	S	Y	8,315	7,560	6,872	Custom	Placed in District and/or School Coordinator kit
Non-scannable Test Books/Scripts			624	18	-	-	262,670	262,670	-		
ELA--PBA	Grade 9 ELA --Test Booklet	8-3/8X10-7/8 booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color + black on cover, one color + black internal	48	1	NS	Y	46,700	46,700		Custom to order	Custom to secure order assignment
	Grade 10 ELA --Test Booklet	" "	48	1	NS	Y	44,113	44,113		"	
	Grade 11 ELA -- Test Booklet	" "	48	1	NS	Y	40,437	40,437		"	
MATH--PBA	Algebra I Test Booklet	8-3/8X10-7/8 booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color + black on cover, one color + black internal	40	2	NS	Y	34,977	34,977		"	
	Algebra II Test Booklet	" "	40	2	NS	Y	33,041	33,041		"	
	Geometry Test Booklet	" "	40	2	NS	Y	30,281	30,281		"	
	Int I Test Booklet	" "	40	1	NS	Y	11,659	11,659		"	
	Int II Test Booklet	" "	40	1	NS	Y	11,011	11,011		"	
	Int III Test Booklet	" "	40	1	NS	Y	10,094	10,094		"	
	Spanish Algebra I Test Booklet	" "	40	1	NS	Y	95	95			
	Spanish Algebra II Test Booklet	" "	40	1	NS	Y	90	90			
	Spanish Geometry Test Booklet	" "	40	1	NS	Y	82	82			
	Spanish Int I Test Booklet	" "	40	1	NS	Y	32	32			
	Spanish Int II Test Booklet	" "	40	1	NS	Y	30	30			
	Spanish Int III Test Booklet	" "	40	1	NS	Y	28	28			
Large Print			624	15	-	-	1,090	1,090	-		
	Large Print Grade 9 ELA Test Book	large print book, 14X17, 150% photo enlargement, black on white 50# stock, coil bind (minimum 18-pt type)	48	1	NS	Y	173	173		1's, Kit	Kits to include instruction memo, corresponding form 1 test book & matching scan A/doc
	Large Print Grade 10 ELA Test Book	" "	48	1	NS	Y	163	163		"	" "
	Large Print Grade 11 ELA Test Book	" "	48	1	NS	Y	149	149		"	" "
	Large Print Algebra I Test Book	" "	40	1	NS	Y	129	129		"	" "
	Large Print Algebra II Test Book	" "	40	1	NS	Y	123	123		"	" "

	Large Print Geometry Test Book	" "	40	1	NS	Y	112	112		"	" "
	Large Print Int I Test Book	" "	40	1	NS	Y	43	43		"	" "
	Large Print Int II Test Book	" "	40	1	NS	Y	41	41		"	" "
	Large Print Int III Test Book	" "	40	1	NS	Y	38	38		"	" "
	Lg Print Spanish Algebra I Test Booklet	" "	40	1	NS	Y	29	29			
	Lg Print Spanish Algebra II Test Booklet	" "	40	1	NS	Y	28	28			
	Lg Print Spanish Geometry Test Booklet	" "	40	1	NS	Y	28	28			
	Lg Print Spanish Int I Test Booklet	" "	40	1	NS	Y	12	12			
	Lg Print Spanish Int II Test Booklet	" "	40	1	NS	Y	11	11			
	Lg Print Spanish Int III Test Booklet	" "	40	1	NS	Y	11	11			
Braille			960	9			98	94	-		
	Braille Grade 9 ELA Test Book	Contracted Braille: 11-1/2x11" Manila 100# stock	120	1	NS	Y	17	17		1's, Kit	Kits to include instructions and matching scan A/doc
	Braille Grade 10 ELA Test Book	" "	120	1	NS	Y	16	16		"	" "
	Braille Grade 11 ELA Test Book	" "	120	1	NS	Y	14	14		"	" "
	Braille Algebra I Test Book	" "	100	1	NS	Y	12	12		"	" "
	Braille Print Algebra II Test Book	" "	100	1	NS	Y	13	12		"	" "
	Braille Print Geometry Test Book	" "	100	1	NS	Y	12	11		"	" "
	Braille Print Int I Test Book	" "	100	1	NS	Y	4	4		"	" "
	Braille Print Int II Test Book	" "	100	1	NS	Y	4	4		"	" "
	Braille Print Int III Test Book	" "	100	1	NS	Y	4	4		"	" "
Other Non-Scannable Printed Materials (Manuals, etc.)			768	6			26,342	23,947	-		
	Test Coordinator Manual (Paper)	8-1/2 x 11 self cover booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	96	1	NS	Y	2,171	1,973		1's	Placed in District and School Coord Kits
	Test Coordinator Manual (CBA)	8-1/2 x 11 self cover booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	96	1	NS	Y	2,171	1,973		1's	Placed in District and School Coord Kits

	Test Admin Manual (Paper) - Gr 9-11 ELA	8-1/2 x 11 self cover booklet, perfect-bound, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	144	1	NS	Y	5,500	5,000		TBD - 5's & 1's or other	
	Test Admin Manual (Paper) Gr 9-11 Mathematics	" "	144	1	NS	Y	5,500	5,000		"	
	Test Admin Manual (CBA) - Gr 9-11 ELA	" "	144	1	NS	Y	5,500	5,000		"	
	Test Admin Manual (CBA) - Gr 9-11 Mathematics	" "	144	1	NS	Y	5,500	5,000		"	
Printed & Other Non-Stock Materials (CD. Cassettes, Etc.)			69	38			271,899	271,899	-		
	Algebra I Reference sheet	8.5 x 11 Single sheet, printed black simplex on 50# white offset, Wrap in 20's	2	1	NS	Y	36,726	36,726		20's or TBD	Will come from the vendor wrapped in 20's or other TBD size
	Algebra II Reference sheet	" "	2	1	NS	Y	34,693	34,693		"	" "
	Geometry Reference sheet	" "	2	1	NS	Y	31,795	31,795		"	" "
	Int I Reference sheet	" "	2	1	NS	Y	12,242	12,242		"	" "
	Int II Reference sheet	" "	2	1	NS	Y	11,562	11,562		"	" "
	Int III Reference sheet	" "	2	1	NS	Y	10,598	10,598		"	" "
	Spanish Algebra I Reference sheet	8.5 x 11 Single sheet, printed black simplex on 50# white offset, Wrap in 20's	2	1	NS	Y	100	100		"	" "
	Spanish Algebra II Reference sheet	" "	2	1	NS	Y	95	95		"	" "
	Spanish Geometry Reference sheet	" "	2	1	NS	Y	86	86		"	" "
	Spanish Int I Reference sheet	" "	2	1	NS	Y	33	33		"	" "
	Spanish Int II Reference sheet	" "	2	1	NS	Y	32	32		"	" "
	Spanish Int III Reference sheet	" "	2	1	NS	Y	29	29		"	" "
LP	Lg Print - Algebra I Reference sheet	Single Lg Print 14X17 sheet, 150% photo enlargement, black on white 50# stock (minimum 18-pt type)	2	1	NS		129	129		1's, Kit	Part of Lg Print kit
LP	Lg Print - Algebra II Reference sheet	" "	2	1	NS		123	123		"	" "
LP	Lg Print - Geometry Reference sheet	" "	2	1	NS		112	112		"	" "
LP	Lg Print - Int I Reference sheet	" "	2	1	NS		43	43		"	" "
LP	Lg Print - Int II Reference sheet	" "	2	1	NS		41	41		"	" "
LP	Lg Print - Int III Reference sheet	" "	2	1	NS		38	38		"	" "
LP	Spanish Lg Print - Algebra I Reference sheet	Single Lg Print 14X17 sheet, 150% photo enlargement, black on white 50# stock (minimum 18-pt type)	2	1	NS		29	29			
LP	Spanish Lg Print - Algebra II Reference sheet	" "	2	1	NS		28	28			
LP	Spanish Lg Print - Geometry Reference sheet	" "	2	1	NS		28	28			
LP	Spanish Lg Print - Int I Reference sheet	" "	2	1	NS		12	12			
LP	Spanish Lg Print - Int II Reference sheet	" "	2	1	NS		11	11			
LP	Spanish Lg Print - Int III Reference sheet	" "	2	1	NS		11	11			
BR	BRAILLE - Algebra I Reference sheet	Contracted Braille: 11-1/2x11" sheet, Manila 100# stock	2	1	NS	Y	12	12		1's, Kit	Part of Braille kit
BR	BRAILLE - Algebra II Reference sheet	" "	2	1	NS	Y	12	12		"	" "
BR	BRAILLE - Geometry Reference sheet	" "	2	1	NS	Y	11	11		"	" "
BR	BRAILLE - Int I Reference sheet	" "	2	1	NS	Y	4	4		"	" "
BR	BRAILLE - Int II Reference sheet	" "	2	1	NS	Y	4	4		"	" "
BR	BRAILLE - Int III Reference sheet	" "	2	1	NS	Y	4	4		"	" "
	Ruler - Grade - 6-HS	Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	NS	N	131,063	131,063		TBD	
	Spanish Ruler - Grade - 6-HS	Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	NS	N	357	357			
LP	LG Print - Ruler - Grade - 6-HS	LP version to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	NS	N	486	486		1's, Kit	Part of Lg Print kit
LP	Spanish LG Print - Ruler - Grade - 6-HS	Spanish LP version to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	NS	N	119	119			

BR	BRAILLE - Ruler - Grade - 6-HS	Brailled to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	NS	N	47	47		1's, Kit	Part of Braille kit
	Large Print Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	NS	N	971	971		"	Insert in Lg Print kit assemby
	Spanish Lg Print instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	NS	N	119	119		"	
	Braille Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	NS	N	94	94		"	Insert in Braille kit assemby
	Read Aloud Kit	Provide uncollated (test book) units for use with students with special accommodations to have the test read aloud in small groups.						-		"	Nothing known related to any additonal materials.
Miscellaneous Materials (Stock)			3	2			305,418	330,654	-		
	Pre-ID Labels	LABEL, PRE-ID, 3900, 3.5X1 7/16 TONER BACK	0.2	1	NS	Y	262,670	262,670		cust	
	Paper Bands	standard 3"x27" brown kraft self-sealing bands	1	1	NS	N	17,511	17,511			
	Stock Colored Return labels	3x5 colored return labels, 2-up, laser compatible	0.5		NS	N	8,412	8,412			
	Stock Outbound Shipping labels	3x5 white shipping labels, 2-up, laser compatible	0.5		NS	N	8,412	8,412			
	UPS Labels	As needed for carrier. Stock	0.5		NS	N	8,412	8,412			

Y1Spr+Y2-4 Winter & Spr HS EOY

Y1 Spr + Y2-4 Winter/Spr Block HS EOY's

1,145 Schools
948 Distribution Points
828 Systems/Districts

	Document Title	Document Description	# Pgs	# of Fms	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
Scannable Materials			150	15	-	-	262,670	262,670	250,000		
ELA--EOY	Grade 9 ELA -- Answer Doc	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	2	1	S	Y	46,700	46,700	44,476	20's, 5's - or other TBD	
	Grade 10 ELA -- Answer Doc	" "	2	1	S	Y	44,113	44,113	42,012	"	
	Grade 11 ELA -- Answer Doc	" "	2	1	S	Y	40,437	40,437	38,512	"	
MATH--EOY	Algebra I Answer Book	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	12	1	S	Y	34,977	34,977	33,311	"	
	Algebra II Answer Book	" "	12	1	S	Y	33,041	33,041	31,468	"	
	Geometry Answer Book	" "	12	1	S	Y	30,281	30,281	28,839	"	
	Int I Answer Book	" "	12	1	S	Y	11,659	11,659	11,104	"	
	Int II Answer Book	" "	12	1	S	Y	11,011	11,011	10,487	"	
	Int III Answer Book	" "	12	1	S	Y	10,094	10,094	9,613	"	
	Spanish Algebra I Answer Book	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	12	1			95	95	48		
	Spanish Algebra II Answer Book	" "	12	1			90	90	45		
	Spanish Geometry Answer Book	" "	12	1			82	82	41		
	Spanish Int I Answer Book	" "	12	1			32	32	16		
	Spanish Int II Answer Book	" "	12	1			30	30	15		
	Spanish Int III Answer Book	" "	12	1			28	28	14		
Scannable Headers			2	1			-	7,560	6,872		
	School/Classroom Header	scannable sheet, 166 (6x6 layout), 9.5x11, mark reflex, black plus one additional color, 60# offset, continuous, lithocode front/back	2	1	S	Y	8,315	7,560	6,872	Custom	Placed in District and/or School Coordinator kit
Non-scannable Test Books/Scripts			528	33	-	-	262,670	262,670	-		
	Grade 9 ELA --Test Booklet	8-3/8X10-7/8 booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson standard is 50# white offset), prints one color + black on cover, one color + black internal	48	5	NS	Y	46,700	46,700		Custom to order	Custom to secure order assignment
	Grade 10 ELA --Test Booklet	" "	48	5	NS	Y	44,113	44,113		"	
	Grade 11 ELA -- Test Booklet	" "	48	5	NS	Y	40,437	40,437		"	
	Algebra I Test Booklet	" "	32	2	NS	Y	34,977	34,977		"	
	Algebra II Test Booklet	" "	32	2	NS	Y	33,041	33,041		"	
	Geometry Test Booklet	" "	32	2	NS	Y	30,281	30,281		"	
	Int I Test Booklet	" "	32	2	NS	Y	11,659	11,659		"	
	Int II Test Booklet	" "	32	2	NS	Y	11,011	11,011		"	
	Int III Test Booklet	" "	32	2	NS	Y	10,094	10,094		"	
	Spanish Algebra I Test Booklet	" "	32	1	NS	Y	95	95			
	Spanish Algebra II Test Booklet	" "	32	1	NS	Y	90	90			
	Spanish Geometry Test Booklet	" "	32	1	NS	Y	82	82			
	Spanish Int I Test Booklet	" "	32	1	NS	Y	32	32			
	Spanish Int II Test Booklet	" "	32	1	NS	Y	30	30			
	Spanish Int III Test Booklet	" "	32	1	NS	Y	28	28			
Large Print			528	15	-	-	1,090	1,090	-		
	Large Print Grade 9 ELA Test Book	large print book, 14X17, 150% photo enlargement, black on white 50# stock, coil bind (minimum 18-pt type)	48	1	NS	Y	173	173		1's, Kit	Kits to include instruction memo, corresponding form 1 test book & matching scan A/doc
	Large Print Grade 10 ELA Test Book	" "	48	1	NS	Y	163	163		"	" "
	Large Print Grade 11 ELA Test Book	" "	48	1	NS	Y	149	149		"	" "
	Large Print Algebra I Test Book	" "	32	1	NS	Y	129	129		"	" "
	Large Print Algebra II Test Book	" "	32	1	NS	Y	123	123		"	" "
	Large Print Geometry Test Book	" "	32	1	NS	Y	112	112		"	" "
	Large Print Int I Test Book	" "	32	1	NS	Y	43	43		"	" "
	Large Print Int II Test Book	" "	32	1	NS	Y	41	41		"	" "
	Large Print Int III Test Book	" "	32	1	NS	Y	38	38		"	" "

	Lg Print Spanish Algebra I Test Booklet	" "	32	1	NS	Y	29	29			
	Lg Print Spanish Algebra II Test Booklet	" "	32	1	NS	Y	28	28			
	Lg Print Spanish Geometry Test Booklet	" "	32	1	NS	Y	28	28			
	Lg Print Spanish Int I Test Booklet	" "	32	1	NS	Y	12	12			
	Lg Print Spanish Int II Test Booklet	" "	32	1	NS	Y	11	11			
	Lg Print Spanish Int III Test Booklet	" "	32	1	NS	Y	11	11			
Braille			840	9			98	94	-		
	Braille Grade 9 ELA Test Book	Contracted Braille: 11-1/2x11" Manila 100# stock	120	1	NS	Y	17	17		1's, Kit	Kits to include instructions and matching scan A/doc
	Braille Grade 10 ELA Test Book	" "	120	1	NS	Y	16	16		"	" "
	Braille Grade 11 ELA Test Book	" "	120	1	NS	Y	14	14		"	" "
	Braille Algebra I Test Book	" "	80	1	NS	Y	12	12		"	" "
	Braille Print Algebra II Test Book	" "	80	1	NS	Y	13	12		"	" "
	Braille Print Geometry Test Book	" "	80	1	NS	Y	12	11		"	" "
	Braille Print Int I Test Book	" "	80	1	NS	Y	4	4		"	" "
	Braille Print Int II Test Book	" "	80	1	NS	Y	4	4		"	" "
	Braille Print Int III Test Book	" "	80	1	NS	Y	4	4		"	" "
Other Non-Scannable Printed Materials (Manuals, etc.)			-	-	-	-	-	-	-		
Printed & Other Non-Stock Materials (CD. Cassettes, Etc.)			68	23			241,999	241,999	-		
	Algebra I Reference sheet	8.5 x 11 Single sheet, printed black simplex on 50# white offset, Wrap in 20's	2	1	NS	Y	36,726	36,726		20's or TBD	Will come from the vendor wrapped in 20's or other TBD size
	Algebra II Reference sheet	" "	2	1	NS	Y	34,693	34,693		"	" "
	Geometry Reference sheet	" "	2	1	NS	Y	31,795	31,795		"	" "
	Int I Reference sheet	" "	2	1	NS	Y	12,242	12,242		"	" "
	Int II Reference sheet	" "	2	1	NS	Y	11,562	11,562		"	" "
	Int III Reference sheet	" "	2	1	NS	Y	10,598	10,598		"	" "
LP	Lg Print - Algebra I Reference sheet	Single Lg Print 14X17 sheet, 150% photo enlargement, black on white 50# stock (minimum 18-pt type)	2	1	NS		129	129		1's, Kit	Part of Lg Print kit
LP	Lg Print - Algebra II Reference sheet	" "	2	1	NS		123	123		"	" "
LP	Lg Print - Geometry Reference sheet	" "	2	1	NS		112	112		"	" "
LP	Lg Print - Int I Reference sheet	" "	2	1	NS		43	43		"	" "
LP	Lg Print - Int II Reference sheet	" "	2	1	NS		41	41		"	" "
LP	Lg Print - Int III Reference sheet	" "	2	1	NS		38	38		"	" "
	Spanish Algebra I Reference sheet	" "	2		NS	Y	29	29			
	Spanish Algebra II Reference sheet	" "	2		NS	Y	28	28			
	Spanish Geometry Reference sheet	" "	2		NS	Y	28	28			
	Spanish Int I Reference sheet	" "	2		NS	Y	12	12			
	Spanish Int II Reference sheet	" "	2		NS	Y	11	11			
	Spanish Int III Reference sheet	Contracted Braille: 11-1/2x11" sheet, Manila 100# stock	2		NS	Y	11	11			
BR	BRAILLE - Algebra I Reference sheet	Contracted Braille: 11-1/2x11" sheet, Manila 100# stock	2	1	NS	Y	12	12		1's, Kit	Part of Braille kit
BR	BRAILLE - Algebra II Reference sheet	" "	2	1	NS	Y	12	12		"	" "
BR	BRAILLE - Geometry Reference sheet	" "	2	1	NS	Y	11	11		"	" "
BR	BRAILLE - Int I Reference sheet	" "	2	1	NS	Y	4	4		"	" "
BR	BRAILLE - Int II Reference sheet	" "	2	1	NS	Y	4	4		"	" "
BR	BRAILLE - Int III Reference sheet	" "	2	1	NS	Y	4	4		"	" "
BR	Spanish Lg Print - Algebra I Reference sheet	" "	2		NS	N	29	29			
BR	Spanish Lg Print - Algebra II Reference sheet	" "	2		NS	N	28	28			
BR	Spanish Lg Print - Geometry Reference sheet	" "	2		NS	N	28	28			
BR	Spanish Lg Print - Int I Reference sheet	" "	2		NS	N	12	12			
BR	Spanish Lg Print - Int II Reference sheet	" "	2		NS	N	11	11			
	Spanish Lg Print - Int III Reference sheet	Qty based on per ruler count - 1/2 and 1/4 inch markings, 0.5 or 0.1 centimeter markings, one per student, disposable-printed cardstock	2		NS	N	11	11			
LP	LG Print - Ruler - Grade 4-5	LP version to support: Qty based on per ruler count - 1/8 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock			NS	N	34,977	34,977		"	" "

LP	LG Print - Ruler - Grade - 6-HS	LP version to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.8	1	NS	N	68,018	68,018		1's, Kit	Part of Lg Print kit
BR	BRAILLE - Ruler - Grade - 6-HS	Brailled to support: Qty based on per ruler count - 1/16 inch and 0.1 centimeter markings minimum, one per student, disposable-printed cardstock	0.4	1	NS	N	47	47		1's, Kit	Part of Braille kit
	(All) Large Print Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	NS	N	476	476		"	Insert in Lg Print kit assemby
	Braille Instructions	8.5 x 11 Single sheet, printed black simplex on 50# white offset	2	1	NS	N	94	94		"	Insert in Braille kit assemby
	Read Aloud Kit	Provide uncollated (test book) units for use with students with special accommodations to have the test read aloud in small groups.						-			Nothing known related to any additonal materials.
Miscellaneous Materials (Stock)			3	2			297,157	314,134	-		
	Pre-ID Labels	LABEL, PRE-ID, 3900, 3.5X1 7/16 TONER BACK	0.2	1	NS	Y	262,670	262,670			
	Paper Bands	standard 3"x27" brown kraft self-sealing bands	1	1	NS	N	17,511	17,511			
	Stock Colored Return labels	3x5 colored return labels, 2-up, laser compatible	0.5		NS	N	5,659	5,659			
	Stock Outbound Shipping labels	3x5 white shipping labels, 2-up, laser compatible	0.5		NS	N	5,659	5,659			
	UPS Labels	As needed for carrier. Stock	0.5		NS	N	5,659	5,659			

Spr 2015 ELA PBA SA FT

2,592 Schools
1,175 Distribution Points
468 Systems/Districts

	Document Title	Document Description	# Pgs	# of Fms	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
Scannable Materials			40	60			612,360	612,360	583,200		
ELA--PBA	Grade 3 ELA -- Consumable Test Booklet	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	24	**36**	S	Y	68,040	68,040	64,800	Custom to order need	Custom to secure order assignment
	Grade 4 ELA -- Answer Book	" "	2	3	S	Y	68,040	68,040	64,800	20's, 5's - or other TBD	
	Grade 5 ELA -- Answer Book	" "	2	3	S	Y	68,040	68,040	64,800	"	
	Grade 6 ELA -- Answer Book	" "	2	3	S	Y	68,040	68,040	64,800	"	
	Grade 7 ELA -- Answer Book	" "	2	3	S	Y	68,040	68,040	64,800	"	
	Grade 8 ELA -- Answer Book	" "	2	3	S	Y	68,040	68,040	64,800	"	
	Grade 9 ELA -- Answer Book	" "	2	3	S	Y	68,040	68,040	64,800	"	
	Grade 10 ELA -- Answer Book	" "	2	3	S	Y	68,040	68,040	64,800	"	
	Grade 11 ELA -- Answer Book	" "	2	3	S	Y	68,040	68,040	64,800	"	
Scannable Headers			2	1			-	2,851	2,592		
	School/Classroom Header	scannable sheet, 166 (6x6 layout), 9.5x11, mark reflex, black plus one additional color, 60# offset, continuous, lithocode front/back	2	1	S	Y	3,136	2,851	2,592	Custom	Placed in District and/or School Coordinator kit
Non-scannable Test Books/Scripts			128	288			544,320	544,320	-		
ELA--PBA	Grade 4 ELA -- Test Booklet	8-3/8X10-7/8 booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color + black on cover, one color	16	**36**	NS	Y	68,040	68,040		20's, 5's - or other TBD	Custom to secure order assignment
	Grade 5 ELA -- Test Booklet	" "	16	**36**	NS	Y	68,040	68,040		"	
	Grade 6 ELA -- Test Booklet	" "	16	**36**	NS	Y	68,040	68,040		"	
	Grade 7 ELA --Test Booklet	" "	16	**36**	NS	Y	68,040	68,040		"	
	Grade 8 ELA -- Test Booklet	" "	16	**36**	NS	Y	68,040	68,040		"	
	Grade 9 ELA --Test Booklet	" "	16	**36**	NS	Y	68,040	68,040		"	
	Grade 10 ELA --Test Booklet	" "	16	**36**	NS	Y	68,040	68,040		"	
	Grade 11 ELA -- Test Booklet	" "	16	**36**	NS	Y	68,040	68,040		"	
Large Print			-	-			-	-	-		
Braille			-	-			-	-	-		
Other Non-Scannable Printed Materials (Manuals, etc.)			1,056	8			58,055	52,777	-		
	Test Coordinator Manual (Paper)	8-1/2 x 11 self cover booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	96	1	NS	Y	3,367	3,060		1's	Placed in District and School Coord Kits
	Test Coordinator Manual (CBA)	" "	96	1	NS	Y	3,367	3,060		1's	Placed in District and School Coord Kits
	Test Admin Manual (Paper) Gr 3-5 ELA	8-1/2 x 11 self cover booklet, perfect-bound, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	144	1	NS	Y	8,554	7,776		TBD - 5's & 1's or other	
	Test Admin Manual (Paper) Gr 6-8 ELA	" "	144	1	NS	Y	8,554	7,776		"	
	Test Admin Manual (Paper) - Gr 9-11 ELA	8-1/2 x 11 self cover booklet, perfect-bound, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	144	1	NS	Y	8,554	7,776		"	
	Test Admin Manual (CBA) Gr 3-5 ELA	" "	144	1	NS	Y	8,554	7,776		"	
	Test Admin Manual (CBA) Gr 6-8 ELA	" "	144	1	NS	Y	8,554	7,776		"	
	Test Admin Manual (CBA) - Gr 9-11 ELA	" "	144	1	NS	Y	8,554	7,776		"	
Non-Printed Non-Stock Materials (CD. Cassettes, Etc.)							-	-	-		
Miscellaneous Materials (Stock)			3	2			678,335	701,132	-		

	Pre-ID Labels	LABEL, PRE-ID, 3900, 3.5X1 7/16 TONER BACK	0.2	1	NS	Y	612,360	612,360				
	Paper Bands	standard 3"x27" brown kraft self-sealing bands	1	1	NS	N	40,824	40,824				
	Stock Colored Return labels	3x5 colored return labels, 2-up, laser compatible	0.5		NS	N	8,384	8,384				
	Stock Outbound Shipping labels	3x5 white shipping labels, 2-up, laser compatible	0.5		NS	N	8,384	8,384				
	UPS Labels	As needed for carrier. Stock	0.5		NS	N	8,384	8,384				

Test Admin Manual (Paper) Gr 6-8 ELA	" "	144	1	NS	Y	7,128	6,480		"	
Test Admin Manual (Paper) - Gr 9-11 ELA	8-1/2 x 11 self cover booklet, perfect-bound, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	144	1	NS	Y	7,128	6,480		"	
Test Admin Manual (CBA) Gr 3-5 ELA	" "	144	1	NS	Y	7,128	6,480		"	
Test Admin Manual (CBA) Gr 6-8 ELA	" "	144	1	NS	Y	7,128	6,480		"	
Test Admin Manual (CBA) - Gr 9-11 ELA	" "	144	1	NS	Y	7,128	6,480		"	
ted Non-Stock Materials (CD. Cassettes, Etc.)						-	-	-		
neous Materials (Stock)		3	2			565,282	588,079	-		
Pre-ID Labels	LABEL, PRE-ID, 3900, 3.5X1 7/16 TONER BACK	0.2	1	NS	Y	510,300	510,300			
Paper Bands	standard 3"x27" brown kraft self-sealing bands	1	1	NS	N	34,020	34,020			
Stock Colored Return labels	3x5 colored return labels, 2-up, laser compatible	0.5		NS	N	6,987	6,987			
Stock Outbound Shipping labels	3x5 white shipping labels, 2-up, laser compatible	0.5		NS	N	6,987	6,987			
UPS Labels	As needed for carrier. Stock	0.5		NS	N	6,987	6,987			

Spr 17 & 18 ELA PBA SA FT

1,728 Schools
783 Distribution Points
312 Systems/Districts

	Document Title	Document Description	# Pgs	# of Fms	Scan/NS	Compose?	Order Qty	Distrib Qty	Process Qty	Wrap Size	Comments
Scannable Materials			40	48			408,240	408,240	388,800		
ELA--PBA	Grade 3 ELA -- Consumable Test Booklet	scannable book, .166 (6x6 layout), 8.5x11, mark reflex, black plus one additional color, 60# offset, saddle-stitched, corner cut, matching lithocode	24	**24**	S	Y	45,360	45,360	43,200	Custom to order need	Custom to secure order assignment
	Grade 4 ELA -- Answer Book	" "	2	3	S	Y	45,360	45,360	43,200	20's, 5's - or other TBD	
	Grade 5 ELA -- Answer Book	" "	2	3	S	Y	45,360	45,360	43,200	"	
	Grade 6 ELA -- Answer Book	" "	2	3	S	Y	45,360	45,360	43,200	"	
	Grade 7 ELA -- Answer Book	" "	2	3	S	Y	45,360	45,360	43,200	"	
	Grade 8 ELA -- Answer Book	" "	2	3	S	Y	45,360	45,360	43,200	"	
	Grade 9 ELA -- Answer Book	" "	2	3	S	Y	45,360	45,360	43,200	"	
	Grade 10 ELA -- Answer Book	" "	2	3	S	Y	45,360	45,360	43,200	"	
	Grade 11 ELA -- Answer Book	" "	2	3	S	Y	45,360	45,360	43,200	"	
Scannable Headers			2	1			-	1,901	1,728		
	School/Classroom Header	scannable sheet, 166 (6x6 layout), 9.5x11, mark reflex, black plus one additional color, 60# offset, continuous, lithocode front/back	2	1	S	Y	2,091	1,901	1,728	Custom	Placed in District and/or School Coordinator kit
Non-scannable Test Books/Scripts			128	192			362,880	362,880	-		
ELA--PBA	Grade 4 ELA -- Test Booklet	8-3/8X10-7/8 booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color + black on cover, one color + black internal	16	**24**	NS	Y	45,360	45,360		20's, 5's - or other TBD	Custom to secure order assignment
	Grade 5 ELA -- Test Booklet	" "	16	**24**	NS	Y	45,360	45,360		"	
	Grade 6 ELA -- Test Booklet	" "	16	**24**	NS	Y	45,360	45,360		"	
	Grade 7 ELA --Test Booklet	" "	16	**24**	NS	Y	45,360	45,360		"	
	Grade 8 ELA -- Test Booklet	" "	16	**24**	NS	Y	45,360	45,360		"	
	Grade 9 ELA --Test Booklet	" "	16	**24**	NS	Y	45,360	45,360		"	
	Grade 10 ELA --Test Booklet	" "	16	**24**	NS	Y	45,360	45,360		"	
	Grade 11 ELA -- Test Booklet	" "	16	**24**	NS	Y	45,360	45,360		"	
Large Print			-	-			-	-	-		
Braille			-	-			-	-	-		
Other Non-Scannable Printed Materials (Manuals, etc.)			1,056	8			38,703	35,185	-		
	Test Coordinator Manual (Paper)	8-1/2 x 11 self cover booklet, saddle stitched, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	96	1	NS	Y	2,244	2,040		1's	Placed in District and School Coord Kits
	Test Coordinator Manual (CBA)	" "	96	1	NS	Y	2,244	2,040		1's	Placed in District and School Coord Kits
	Test Admin Manual (Paper) Gr 3-5 ELA	8-1/2 x 11 self cover booklet, perfect-bound, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	144	1	NS	Y	5,702	5,184		TBD - 5's & 1's or other	
	Test Admin Manual (Paper) Gr 6-8 ELA	" "	144	1	NS	Y	5,702	5,184		"	
	Test Admin Manual (Paper) - Gr 9-11 ELA	8-1/2 x 11 self cover booklet, perfect-bound, 45# white cavalier or approved equivalent paper (Pearson's standard is 50# white offset), prints one color plus black on cover, black all interior	144	1	NS	Y	5,702	5,184		"	
	Test Admin Manual (CBA) Gr 3-5 ELA	" "	144	1	NS	Y	5,702	5,184		"	
	Test Admin Manual (CBA) Gr 6-8 ELA	" "	144	1	NS	Y	5,702	5,184		"	
	Test Admin Manual (CBA) - Gr 9-11 ELA	" "	144	1	NS	Y	5,702	5,184		"	
Non-Printed Non-Stock Materials (CD. Cassettes, Etc.)							-	-	-		
Miscellaneous Materials (Stock)			3	2			452,229	475,026	-		

	Pre-ID Labels	LABEL, PRE-ID, 3900, 3.5X1 7/16 TONER BACK	0.2	1	NS	Y	408,240	408,240				
	Paper Bands	standard 3"x27" brown kraft self-sealing bands	1	1	NS	N	27,216	27,216				
	Stock Colored Return labels	3x5 colored return labels, 2-up, laser compatible	0.5		NS	N	5,591	5,591				
	Stock Outbound Shipping labels	3x5 white shipping labels, 2-up, laser compatible	0.5		NS	N	5,591	5,591				
	UPS Labels	As needed for carrier. Stock	0.5		NS	N	5,591	5,591				

PARCC Operational: Paper-Based Student Reports

HS Winter Rpts

P/P Students x2 (PPT+intended CBA students)	Single-sheets (2-pg) Rpts x2	Outbound pgs + Print Images Duplex print
500,000	**1,000,000**	**2,000,000**
Pre-printed report form (single-sheet 2-pg) order qty	**1,050,000**	

2,291	Schools
1,896	Distribution Points
1,656	**Systems/Districts**

Custom pre-print form - 2-color + black, single sheet, finished 8.5 x 11
Student ISR's ONLY - assume 2 copies per/student
Includes CBA students (Y1 ONLY - CBA students managed via PPT delivery)

Gr 3-8 Spring Rpts

P/P Students x2 (PPT+CBA students)	4-pg folders Rpts x2	Outbound pgs + Print Images Duplex print
5,640,224	11,280,448	**45,121,792**
Pre-printed report form (4-pg folder) order nct	**11,844,470**	

24,920	**Schools**
11,294	**Distribution Points**
4,504	**Systems/Districts**

Gr 9-11 Spring Rpts

P/P Students x2 (PPT+CBA students)	Single-sheets (2-pg) Rpts x2 (CBA students)	Outbound pgs + Print Images Duplex print
859,776	1,719,552	**3,439,104**
Pre-printed report form (single-sheet 2-pg) order qty	**1,805,530**	
Total reports	**13,000,000**	**48,560,896**

Gr 3-8 Custom pre-print form - 2-color + black, 4-pg (11x17) folder, folded-finish 8.5x11
HS Custom pre-print form - 2-color + black, single sheet, finished 8.5 x 11
Student ISR's ONLY - assume 2 copies per/student
Includes CBA students

HS Spring Rpts

P/P Students x2 (PPT+CBA students)	Single-sheets (2-pg) Rpts x2 (CBA students)	Outbound pgs + Print Images Duplex print
500,000	1,000,000	2,000,000
Pre-printed report form (single-sheet 2-pg) order qty	**1,050,000**	

2,291	**Schools**
1,896	**Distribution Points**
1,656	**Systems/Districts**

Custom pre-print form - 2-color + black, single sheet, finished 8.5 x 11
Student ISR's ONLY - assume 2 copies per/student
Includes CBA students