

VII. Tasks for Graduation Competency Assessments

VII.A. MODULAR FORMAT



DRC understands PDE's requirement that the GCAs be developed and administered in a modular format to allow students who do not achieve proficiency in a specific content area to retake that content area module, rather than the entire GCA assessment. All test designs outlined in this proposal reflect the modular format. Standards will be set at the overall test level to produce four performance levels. Additionally, a non-proficient/proficient (i.e., non-pass/pass) standard (cutscore) will be produced per module, to ensure that required retesting requirements are met. Please refer to *Subheading VII.B.2* for information about operational test designs. For detailed information about DRC's proposed standard setting approach, see *Subheading VII.J.8*.

VII.B.2 for information about operational test designs. For detailed information about DRC's proposed standard setting approach, see *Subheading VII.J.8*.

VII.B. ITEM DEVELOPMENT

DRC and WestEd are pleased to propose an item development plan that will ensure that the Commonwealth assesses student, school, and district performance with validity, reliability, and efficiency. Over the past five years, we have developed the staff, knowledge, expertise, and processes necessary to address effectively all key components of the PSSA's item development needs—alignment to content standards; assessment anchors and eligible content; universal design; bias, sensitivity, fairness issues—and adherence to item specifications and approved style and format. The partnership of our two companies represents the combined wisdom and track record of two assessment industry leaders. We look forward to our continued partnership as we work to support PDE in the development of the Graduation Competency Assessment (GCA) and to work with Pennsylvania educators in this phase of assessment development.

The DRC and WestEd development team is uniquely qualified to provide the item development services as outlined in the RFP. In our previous work on the PSSA, we have established a consistent record of reliability and quality. We have delivered error-free items and test forms that have been integral to supporting Pennsylvania's assessment and reform efforts; we are committed to continue providing this same level of quality service. DRC and WestEd offer PDE many strengths and advantages through our continued partnership as Pennsylvania's item and test development contractors. The DRC team will develop the Biology, Chemistry, and English Composition GCA. Literature, Algebra I, Algebra II, Geometry, World History, U.S. History, and Civics and Government will be developed by WestEd's development team.

Knowledge of Pennsylvania's Standards and Anchors

Based on our acquired knowledge and experience in Pennsylvania, the DRC and WestEd team is grounded in a solid understanding of the standards that form the foundation of the Pennsylvania's assessments. We have worked closely with PDE as decisions about PSSA assessable anchors, eligible content, and cognitive complexity have been made over the life of the contract. In so doing, we have worked together with PDE as the Commonwealth has added rigor to the assessment through increased demands for cognitive complexity.

In each of the past five years, we have successfully worked with PDE to produce items and assessments that align to the identified assessable anchors for reading, mathematics, and science and the standards for writing that assess student knowledge and skill at the desired level of cognitive complexity. In recent years, Pennsylvania has classified items using Norman Webb's depth-of-knowledge framework (Webb, N.L. 1997, 2007), and DRC and WestEd staffs have extensive knowledge and experience using Webb's Depth of Knowledge to classify cognitive complexity.¹ In determining the depth of knowledge level for each item, the content specialists at DRC and WestEd have worked closely with PDE staff in order to internalize and implement PDE's definition for the overall cognitive complexity by clarifying objectives and developing strategies to expand the depth of knowledge of the item pool. Our team of item developers is well positioned to continue to meet the Commonwealth's goals for complexity within the GCA.

DRC's and WestEd's thorough understanding of the Pennsylvania assessment, the assessment anchors, the content standards, and the desired cognitive complexity of the GCA, places us in a unique position to respond to any future changes in the assessment anchors or other parameters of the development. Going forward, we face no "learning curve" in developing such an understanding because our item development team is conversant in all aspects of the content that is at the heart of the GCA for all content areas: reading, mathematics, writing, science, and social studies.

Understanding of Pennsylvania Item Specifications

DRC and WestEd will work closely with PDE to specify features and parameters of GCA items. Per the RFP, DRC will use PSSA item specifications for GCA item development and modify them, if necessary. Due to the new social studies GCAs, item specifications for the development of World History, U.S. History, and Civics and Government will be developed by WestEd, and reviewed with PDE and DRC via conference call as stated in the RFP.

Whether the issue relates to acceptable choices for point of view in a reading item, acceptable fractions to use in a mathematics item, graphics parameters for a diagram, acceptable topics or themes for science scenarios, or the understanding of subjects or topics that will engage student interest in responding to writing

¹ We are also well familiar with the other methodologies for classifying cognitive complexity that are widely in use, including those developed by Bloom et al. and Porter.

prompts, we have internalized item specifications and that enables us to efficiently develop items that are consistent with the expectations of PDE.

Knowledge of Pennsylvania Style Guide

Knowledge of and consistent implementation of the Style Guide is critical to the development of items that reflect a uniform presentation of content that support the reliability and validity of the assessment. A consistent style helps to remove irrelevant elements from the assessment. DRC will work with PDE to create the Pennsylvania GCA-specific style guide which will be consistent to the PSSA Style Guide. Our test development staff and lead test development illustrators and designers will meet with PDE to determine a specific style guide that includes Pennsylvania-specific conventions for spelling, grammar, syntax, word choice, etc. for the GCA. This guide will be used along with the item specifications as a set of guidelines for item/test development. Style guide considerations for the development of the social studies GCA will involve input from WestEd. After approval from PDE, the guide will be used continuously during materials development to incorporate all client stylistic preferences.

Our adherence to PDE-approved style extends to the wording and format of the items, the specifications for graphics, and the final presentation of items within the test booklets. This is reflected in overall consistency of phrasing, development of answer choices that do not stand out such that all are plausible and logical, consistent labeling of figures and graphics, and the use of a common font style and size throughout the item pool. This knowledge is shared by all of our item writers, editors, and publishing staff, including item banking staff members.

The item authoring and management tool (IDEAS) is customizable for Pennsylvania so that item writers and editors are always seeing a close approximation of the item will appear in a test booklet. Likewise, IDEAS prints items for content reviews and bias, sensitivity, and fairness reviews consistent with the Style Guide to ensure that educators are evaluating the items as students would encounter them in the test situation. By implementing the Style Guide throughout all stages of the item and test development process, efficiencies are achieved during production because items do not have to be edited for style or format once they have been placed in forms. Edits at that point in the development process lead to increased production costs and significantly increase the risk of introducing errors into the forms. Additional information concerning our IDEAS item banking system can be found later in our item development section.

Knowledge of Universal Design Principles

Our Pennsylvania-specific knowledge and expertise is further complemented by our knowledge of the Principles of Universal Design. We share the commitment with PDE to develop a fair test that provides an accurate measure of what all assessed students know and can do without compromising reliability or validity. In so doing, members of our leadership item-development team have received direct training from the National Center on Educational Outcomes (NCEO);

therefore, we carefully employ the Principles of Universal Design throughout all stages of both the item development process and the test development process. In addition, DRC's Director of Alternate Assessment provides an independent review of all items. In this review, items, graphics, etc., are reviewed for adherence to the Principles of Universal Design. Our Director of Alternate Assessment has also been trained by NCEO and the American Printing House for the Blind.

The elements of universal design that characterize sound assessment practice are incorporated throughout our development practices and processes. The Principles of Universal Design were created to ensure accessible environments for all people through equitable use, simple and intuitive design, effective communication, tolerance for variability, and minimal fatigue. Their application is defended by research that links the Principles of Universal Design to higher performance for all students.

DRC and WestEd have found that explicit universal design considerations are critical because they provide a systematic means for developing assessments in which the context for testing, user diversity, and equitable access are examined at each step of the process. All phases of the DRC and WestEd item development and subsequent test development cycles, reflect the integration of universal design principles with sound measurement theory, current research, and best practices in assessment.

Utilizing the National Center for Educational Outcomes published guidelines (Thompson, et al., 2002) for universal design and the training we received from NCEO, we have incorporated these principles in both the development of items and the subsequent layout of test forms. All item developers, editors, graphic artists, and publications experts are trained in applying universal design principles. Our current item writing and editing practices include the following:

- Using consistent naming and graphics conventions.
- Replacing low-frequency words with simple common words.
- Avoiding irregularly-spelled words, words with ambiguous or multiple meanings, technical terms unless defined and integral to meaning, and concepts with multiple names, symbols, or representations.
- Ensuring clarity of noun-pronoun relationships.
- Simplifying keys and legends.

We recognize that the Commonwealth has a legal and ethical obligation to ensure that Pennsylvania's assessments are accessible and fair to all students. Implementation of universal design principles will contribute to participation by the widest range of students in the assessment program and provide support for the validity of inferences about levels of student performance. By focusing attention on inclusive design principles and providing for a full range of test performances during item development, item quality will be improved. As these

items are intended to complement current pedagogical practices, test results are likely to give a more accurate picture of what students really know and can do in key content areas.

Educators now agree that adherence to universal design principles may offer a valid alternative to the traditional methods of meeting individuals' needs. Current research suggests that the development of universally designed assessments may be a feasible, effective alternative to the use of some accommodations and that the implementation of proactive design principles can reduce the need for some kinds of assistive technology devices and assistive services by building in accommodations for individuals with disabilities before rather than after test production (National Research Council, 2001; Assistive Technology Act of 1998; Scott, et al., 2003). We are especially knowledgeable of applying the Principles of Universal Design not only when items are initially developed but also when graphics are created. Further information concerning our approach to applying the Principles of Universal Design throughout all stages of the development of graphics can be found in *Subheading VII.B., Item Development*.

Knowledge of Considerations for Bias, Sensitivity, and Fairness

Just as considerations for universal design are built into our item development processes and test development processes, the same can be said for our attention to issues related to bias and sensitivity. The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) emphasize the importance of evaluating the language, symbols, words, phrases, and content used in items that could be considered biased or offensive to members of racial, ethnic, gender, or other groups.

DRC and WestEd have established guidelines for bias, fairness, and sensitivity that are key elements of the training for all item and test development staff. The areas addressed by this training include stereotyping and issues related to gender, socioeconomics/class, religion, regional/geographic differences, disability status, and biases against particular age groups. In addition, DRC has catalogued topics that should be avoided and purposefully maintains balance in gender and ethnic emphasis within the pool of available items and passages. Please see *Appendix 2* for DRC's *Guidelines for Bias, Fairness, and Sensitivity Issues*. More information about Bias, Fairness, and Sensitivity training is discussed later in this proposal.

The DRC–WestEd Partnership

DRC and WestEd have forged an effective partnership to produce the highest quality items for the GCA. The strength of the partnership rests on four pillars: an effective staffing plan, strong content knowledge and expertise, a commitment to quality procedures for processing items, and a commitment to work continuously to internalize PDE's vision of how GCA items should be written to best measure the standards.

The DRC and WestEd item development team will consist of senior-level item and test development management staff, content-area item and test developers and

item writers, content-area editorial specialists, bias/fairness specialists, copy editors and proofreaders, content-area graphic artists and publishing experts, item banking experts, permissions editors, and performance assessment leaders. Members of the team have degrees in education, measurement and research, curriculum and instruction, and/or related fields. In addition, the majority of the item development team members who will be writing the items and/or overseeing the process have experience developing items and tests for the PSSA.

Our commitment to a successful implementation of this program is important to our item development team members, and this commitment will be apparent as we respond to PDE's needs and requests throughout all stages of the program. Establishment of a close working relationship with PDE has always been of extreme importance to our item development team members. We understand that all tasks associated with specific goals and objectives need to be responded to quickly. PDE can be assured that activities outlined in our proposal for item development will have the highest priority at DRC and at WestEd.

Effective Staffing Plan

DRC and WestEd's staffing plan ensures that high-quality items are developed and that project communication is smooth and efficient. The Test Development Manager at DRC and the Test Development Manager at WestEd execute the day-to-day project management for item development and test development tasks. Tasks include, but are not limited to the following: project scheduling, coordination with content directors in the oversight of content leads and coordinators, facilitation of weekly meetings and conference calls among staff, budget management, management of schedule changes and any other changes to the scope of work, and communication with PDE.

Our **content directors** supervise the content leads and provide guidance for overall staffing and content. The directors provide oversight and auditing of the final review of all items before delivery to PDE to ensure item quality and accuracy of content. **Pennsylvania content leads** (WestEd literature, mathematics, and social studies; DRC English Composition and science) work together with the **content directors** to oversee the training and development of item writers and editors. Content leads work closely with other staff, such as content directors, graphic artists, and publishing experts to ensure item accuracy and alignment with Pennsylvania standards and style. The content leads also are senior-level editors of items who provide the final review of every developed item prior to delivery. **Item editors** shape the items produced by writers into a more polished product. The content expertise and assessment knowledge of the editors ensures that the items conform to the rigorous content and style guidelines required of GCA items. Both DRC and WestEd employ multiple rounds of editing consistent with advancing levels of proficiency of the editorial staff. The initial development of items comes from the **item writers**, who possess expertise in the content to which they are assigned.

Supporting these content professionals is a well-trained team of **test coordinators, graphic artists, item banking support staff, proofreaders, and permissions editors**. The test coordinator maintains the project calendar for a content area. The test coordinator also monitors the completion of writing, editing, and proofreading assignments. In this role, the test coordinator is in close communication with the content lead regarding the flow of items throughout the writing and editorial process. The graphic artists create the graphics required for items following exacting specifications to ensure content integrity and adherence to the specifications outlined in the Style Guide. The content lead works closely with the **item banking experts** since publishing style guidelines are included in the item banking system. Item banking experts provide the necessary training for Pennsylvania style. Similarly, our **proofreaders** receive training on Pennsylvania style such that items delivered to PDE for review reflect the expected style. The **permissions editors** secure permissions for copyrighted material in a timely manner.

Detailed résumés for our team members can be found in *Appendix I*. Please see *Section E, Personnel*, for our complete approach to staffing for the GCA Program.

Strong Content Knowledge

DRC and WestEd’s combined philosophy of standards-based item development involves the importance of making sure that our thinking is in alignment with the vision of PDE, and we embrace the philosophy that no item, passage, prompt, or scenario, should be written, reviewed, or edited until our item developers have a thorough understanding of PDE’s vision of the GCA and the goals of each program for students in the Commonwealth of Pennsylvania.

The DRC and WestEd team have a deep knowledge of the content areas in which we develop items and of the assessment development practices that support high quality. Our assessment professionals, many of whom are former classroom teachers, also understand how students think and approach test items. Such knowledge helps to ensure that the items will function as part of a valid, reliable assessment to produce accurate measurement of each student’s content knowledge. For example, in a multiple-choice item, understanding how students approach the item can help to eliminate problematic or confusing distractors. In other cases, the close familiarity with the grade level will ensure that the wording in items is grade level appropriate. Clear answer options and grade level appropriate wording are just two parts of our broader understanding of the technical qualities that make a good item. In addition, while all our item developers have a background in national standards, they have immersed themselves in Pennsylvania academic standards, so they know not only what makes a good item, but also what makes a good *Pennsylvania* item.

Our item development efforts are also supported by DRC’s **performance assessment staff**, which reviews all open-ended items and writing prompts. Not only do these staff members know what makes a good open-ended item and writing prompt, but they also know Pennsylvania students and how Pennsylvania

students respond to open-ended items and writing prompts. Please see the heading titled *Select and Train Content Area Item Writers* for each content area of our proposal for more information about the preparation of the item development team.

Item Writer Qualifications

DRC and WestEd’s item and test development team propose to supplement our internal staff content area writers with writers who have experience writing items for PSSA and selected writers from across the country who have written for our companies in the past. Much like our internal staff, many of the external content area item writers have past experience writing items for the PSSA. All content area item writers from both DRC and WestEd are experienced writers, teachers, or former teachers who have a great deal of specialized knowledge in the subject area of their expertise. In selecting the pool of potential content area item writers for reading, writing, mathematics, science, or social studies, care will be taken to select the best writers—qualified individuals who possess both content expertise and good technical writing skills.

In order to select the best writers, a screening process is used. The qualifications we use to select item writers include the following:

- A bachelor’s degree or higher in Algebra I, Algebra II, Geometry, Literature, English Composition, Biology, Chemistry, World History, U.S. History, Civics and Government, curriculum and instruction, and/or related field.
- In-depth understanding and knowledge of the special considerations involving the following: the writing of standards-based multiple-choice items, including writing distractor rationales for each answer option for mathematics and other assessments as requested by PDE; an understanding of depth-of-knowledge levels, estimated difficulty levels, grade-level appropriateness, readability, bias considerations, etc; the development of open-ended items, including developing item-specific scoring guidelines for each item; and the writing of unique, independent, items for passages that do not clue or clang.
- For the English Composition GCA, it is vital that students are engaged in the writing process in a meaningful way; therefore, experience in constructing writing prompts for standards-based tests and/or developing stimulus-based passages with multiple-choice items is necessary.
- In-depth understanding and knowledge of the special considerations involving the construction of writing prompts or scenarios including an understanding of writing mode or scenario purpose, appropriate prompt and scenario scaffolding, composition theory and principles, as well as, difficulty levels, grade-level appropriateness, readability, scorability, bias considerations, etc. DRC understands that prompts and scenarios must be of the highest quality.

- Participation in the assessment-specific training workshop.
- Any other requirements as provided to us by PDE.

Applying the Principles of Universal Design to Graphics

DRC and WestEd understand the challenges of creating items for students with visual impairments. Accordingly, if PDE approves, we will also format the items and create graphics to allow interpretation in sign language, publication in Braille, accommodation for use of assistive technology, and translation into other languages. A universally designed assessment appears clean and organized to the test taker. Our goal is transparency in process, so that the assessment is all-inclusive in its design without compromising the constructs of the material being assessed.

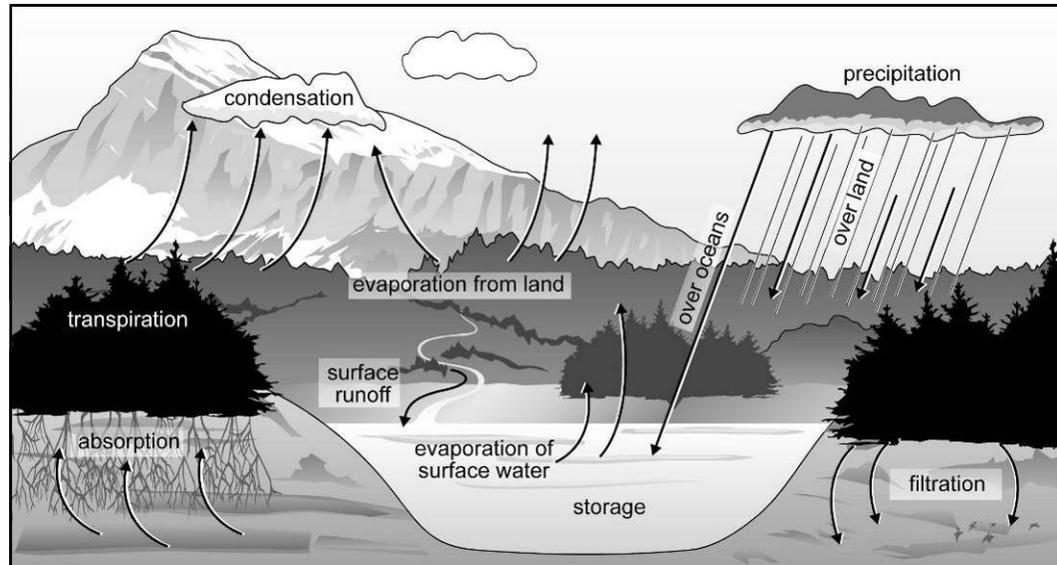
As items are developed, our item development team seeks to ensure the technical quality of all art/graphics produced by our dedicated staff of skilled graphic artists. Each graphic, table, and piece of artwork will depict all relevant or necessary information required by the passage and/or item. Labels will be typeset, appropriate, consistent, and executed according to the established criteria as required by the PDE style guide and be of professional quality.

We are pleased to provide PDE with experienced graphic artists who have been developing graphics for the PSSA since 2003. Our team of graphic artists will lend this experience to the GCA. They have developed a process that integrates the creative aspects of writing item text with the production of visual components that complement the text. During the item-writing stages of the process, the goal is to produce items that are fair and valid for all students, whether a visual component is necessary or not. The desired outcome of this creative process is a link in which text and art are seamless.

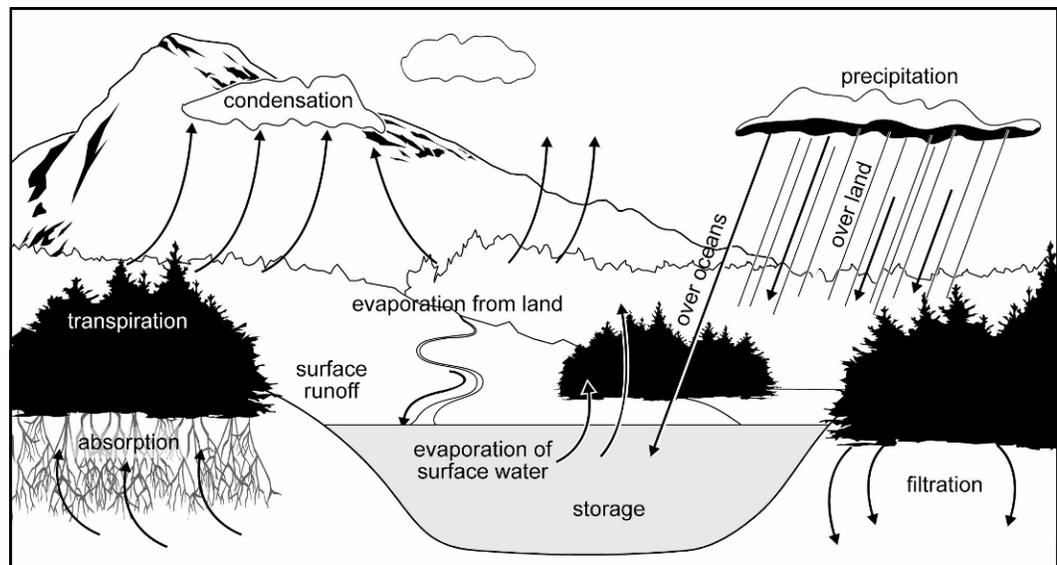
The first step in the process is to ensure that the visual component is necessary. This is especially true in the development of mathematics and science assessments, because graphics are essential elements related to quality mathematics and science items and assessments. Our internal item and test development graphic artists work hand-in-hand with our item development team members and will join the creative process as soon as a visual component of an item is requested. This request will initiate an effort to determine the appropriate balance between art and words for presenting the item. A key consideration at this point is the complexity and scope of the graphics versus the reading load of the text. Complex graphics immediately present a Universal Design challenge to the graphic artist and content specialists. DRC's philosophy involves challenging the content specialists and the graphic artists to develop graphics that are as clear and precise as possible.

The figures below show an original graphic and how item writers, content specialists, and graphic artists might work to ensure that Universal Design is carefully considered for all GCA items. By showing examples of graphics that do not adequately consider the principles of Universal Design during the item-

writing training workshops, DRC and WestEd will challenge our writers to provide feedback as to how graphics can be altered to adhere to the Principles of Universal Design, while maintaining alignment to the *Pennsylvania's Academic Standards*.



Example of an Original Graphic



Example of Modified Graphic with Universal Design Principles Applied

We will strive to create graphics as simple as possible. We will advise graphic artists to restrict the use of shadings, shadows, and colors, as well as fields or indices that are not required for the specific item. This allows the student to focus more on the information needed to answer the question, rather than on reading the graphic.

Item Development Process

This section presents our item development process for the Graduation Competency Assessment (GCA), including the tasks required for the development of items, writing prompts, passages, and scenarios. Our model of development process for item development follows the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) since items are developed to reflect the range of cognitive ability inherent in the standards, resulting in reliable and instructionally valid tests. In addition, our item development work plan also adheres to the Principles of Universal Design, and it reflects that we clearly understand how items and tests must lend themselves to accessibility by diverse groups of students and function appropriately across a broad range of test administration accommodations.

The item development process provides detail associated with each major task in the item development process from the initial meeting with PDE to confirm our understanding of the item development requirements of the program to the final approval of items, writing passages, prompts, and scenarios by PDE for potential field testing. The item development process is outlined below and is also discussed throughout *Subheading VII.B* of our proposal, to follow the outline of the RFP. We have also included a detailed process chart in *Appendix 2*, which follows the process sequentially. This process will be revised on a yearly basis to meet the specific needs and requirements of the program and PDE staff.

Item Development Process

- Meet with PDE: item development planning meeting
- Select and train item writers
- Development of items, passages, prompts, scenarios including graphics
- Review and revise items prior to submission to PDE (internal editing checks)
- Prepare items for review by PDE
- Prepare all materials for content item and bias review
- Support PDE with the content item review committee meeting process
- Prepare written summary reports of the content item review meetings
- Support PDE with the bias, sensitivity, fairness review process
- Prepare written summary reports of the bias, sensitivity, fairness review meetings
- Revise items and conduct internal review process
- Select items for field testing; submit selections to PDE for approval

Item Development Planning Meeting

Prior to beginning the item development process, we propose to meet with PDE to confirm our team's understanding of the item development needs of the program for each year (e.g., number of items per a given academic standard or anchor, number of multiple-choice items and open-ended items per each reading passage, number of items per each writing revising and editing stimulus-based passage, science scenarios, and social studies items), including the plan for the standalone field test in Spring 2010 and number of items to be embedded in field test item positions each operational administration. The meeting will also include a review of all steps in the PDE item approval process and include an overview of our item writing training materials; bias, sensitivity, and fairness guidelines; universal design guidelines; etc., in order to receive feedback from PDE. We will also receive feedback from PDE as to whether there will be any changes to style, item specifications, and target complexity levels.

Prior to meeting with PDE and prior to beginning the item development process for each year's cycle of development, DRC will develop a GCA item bank of items, passages, prompts, and scenarios. The item bank will be developed using the following criteria:

- Match to Pennsylvania Academic Standards for reading, writing, mathematics, science, and social studies.
- Cognitive level alignment.
- Depth-of-knowledge alignment.
- General technical quality, including adherence to industry-standard quality guidelines, along with adherence to PDE's guidelines.
- Adherence to the psychometric guidelines of the GCA.
- Adherence to Principles of Universal Design.
- Freedom from issues of bias, fairness, and sensitivity.
- Other criteria as required by PDE.

In addition, DRC will generate a preliminary plan that will include an overview of the creation, review, and approval processes, as well as a projected schedule for development of items, passages, prompts, and scenarios, including the format for development for PDE review and subsequent committee reviews by Pennsylvania educators.

We will then provide the information to PDE with our recommendation as to how best to target the item writing and prompt development each year to meet the need for the number of items, passages, prompts, and scenarios, as required by the RFP. Per Appendix A of the RFP, all items will be developed to depth-of-knowledge levels two and three.

Select and Train Content Area Item Writers

For all items, our item-development team will prepare item writing training materials and train item writers. We welcome the opportunity for PDE staff to attend and monitor internal item-writing training should PDE request. Costs for attendance could be discussed upon contract award. After training, our item-development team will assign writers to begin the process of writing items, passages, prompts, and scenarios.

A strong GCA program is built upon sound assessment items that are instructionally sensitive and that align to the standards—all of which begin with the selection and training of knowledgeable and experienced content area item writers who have training in developing items that are accessible to all students.

Instructional sensitivity is particularly important given the items to be released to the public through the Item and Scoring Samplers. The released items are an important vehicle for demonstrating the expected knowledge and skills as outlined by the assessment anchors and the standards. Further, instructional leaders and classroom teachers must be able to use the released items to inform and guide instruction. DRC and WestEd item development teams are committed to providing PDE with items that provide the optimal match to the *Pennsylvania's Academic Standards* for Reading, Mathematics, Writing, Science, and Social Studies and that establish clear, focused expectations for grade-level performance by tightly defining the rigor required for grade-level proficiency. This commitment to excellence is reinforced at all steps in the item writer selection and training process. These assurances will also be met through the annual review and refinement of the item specifications and adherence to our item and test development processes, which are designed to ensure alignment.

In the sections below, we outline our proposed approach to the development of items and passages for each content area. This includes additional information about developing the style guide, and the selection and training of content area item writers.

Development of Items, Passages, Prompts, Scenarios Including Graphics

Content area item writers will develop items aligned to the appropriate *Assessment Anchor Content Standard* or the *Pennsylvania's Academic Standard*. Items and prompts will meet PDE-approved style guidelines and item specifications. Items, including all associated passages, stimulus prompts, graphics, and scenarios, will be entered into IDEAS, our item banking system, following all PDE-approved item characteristic requirements. Distractor analysis; assessment anchor or academic standard code for what each item measures; depth-of-knowledge; estimated item difficulty; answer key; scoring guidelines for any open-ended item and writing prompt; and any associated artwork or graphic will also be entered into the system.

Mathematics (Algebra I, Algebra II, and Geometry)

Training Activities for Mathematics Item Writers

DRC and WestEd believe that providing a comprehensive training program designed specifically for potential content area item writers to gain knowledge of the item writing process is a vital component in ensuring quality control in the item development process. Before the first item of a development cycle is written, all levels of the item and test development staff will participate in training. For the development of Algebra I, Algebra II, and Geometry, the training will be organized and conducted by Mr. R. Scott Firkins, the mathematics content lead at WestEd. The training will include a general overview of Pennsylvania items and style and any development guidelines based on decisions of PDE. This training is designed to provide an orientation to the task, specifications, and style in advance of development. This initial training fits into a larger network of staff training achieved through ongoing feedback with PDE throughout the development process and immediate communication about any changes to development specifications. In addition, role-specific training is delivered as described below.

Training Mathematics Item Writers

WestEd propose the use of a combination of both in-house and contract item writers for GCA mathematics item writing. Item writers receive training in the understanding of the Pennsylvania anchors, the expectations for alignment to the Pennsylvania anchors, the item specifications, and the Pennsylvania style guide, best practices for writing technically sound items, searching for and documenting authentic data to use in items, and using IDEAS, the item banking system. All writers will be trained at the start of the year, even if they have worked on the project previously. Item writers and editors will receive a training manual at the start of each development cycle. DRC's training manual, *Item Writer Manual*, can be found in *Appendix 2*. In addition, the content lead will hold regular meetings with content area item writers to give feedback and discuss common issues that arise during the item-writing process.

During the training, examples of items will be provided. It has been the experience of WestEd content-area item and test development staff that writers need to be aware of the reasons why items might be rejected. Providing sample items during the item writing training workshop will allow writers to have a better understanding of what makes a high-quality and technically sound test item.

The writers on the mathematics development team are specialists in Algebra I, Algebra II, and Geometry. Assignments are made so that each assessable standard reflects the thinking of multiple writers; that is, unless the item order calls for only one or two items in a particular eligible content, no single writer will develop all the items for an academic standard.

Writing Mathematics Items

Once a content area writer has been trained and an item assignment has been received, the writers will follow the item writer/editor guidelines and GCA style guide when writing their items. The following fields of information are completed by the content area item writer within DRC's item banking system; IDEAS (please refer to the heading labeled *Item banking System*, presented later in this section, for more information regarding IDEAS):

- Depth of Knowledge
- Estimated difficulty
- Focus
- Calculator usage (determined by the anchor but flagged by the writer)
- Graphics flag (item contains graphics)
- Stem
- Answer choices

WestEd will develop mathematics items for Algebra I, Algebra II, and Geometry. Mathematics items will be developed to meet Pennsylvania specifications. Each mathematics item will have its associated metadata, including: a unique item number, content area, reporting category, academic standard, sub-anchor, eligible content, focus, item type, eligibility for the calculator section, Webb's depth of knowledge, estimated difficulty, and answer key. The trained content experts developing the items will ensure that items are grade-level appropriate. Items will be entered into IDEAS, our item banking system.

- Answer choice rationales
- Correct answer
- Exemplar response
- Source (if necessary)
- Writer notes (optional)

Using the item-authoring tools within IDEAS, the item writer will input the item's text, answer options, and distractor rationales directly into the system. The writer can enter expressions, formulas, equations, and symbols via a MathFlow™ plug-in that works seamlessly with IDEAS. For items requiring graphics, the writer enters a description of the graphic in a notes field and checks a box indicating that a graphic needs to be created. The writer may sketch out the idea for the graphic and send it to the desktop publisher via facsimile or via scanning-to-email technology within DRC's secure email system. In addition, all contract item writers will be required to sign a Confidentiality Statement in which they agree to treat all materials and communication related to item development as confidential.

Science (Biology, Chemistry)

Training Activities for Science Item Writers

All levels of the item and test development staff will participate in training before the development cycle begins. This training will be organized and conducted by Mr. David Durette, the science content lead, at DRC. The training will follow the same process as described in mathematics, but it will also include special consideration of the development of science scenarios. In addition, the set of items developed for a scenario will span depth-of-knowledge levels. It is critical that content area item writers and editors acknowledge and maintain the depth-of-knowledge balance throughout the item writing and editing process.

Training Science Item Writers

Each content area writer will also be asked to pay careful attention to the readability of each item and scenario and check to ensure that the focus is upon the concepts, not upon reading. The goal is for each writer to write items and scenarios that are, to the greatest degree possible, independent of the assessment of reading; however, writers will receive instruction concerning grade-appropriate *science* vocabulary.

Writing Science Items

Science content area item writers will input items, stems, answer options, and distractor rationales directly into IDEAS, our item banking system. Other pertinent information associated with the item (e.g., depth of knowledge, estimated difficulty, focus) will also be entered.

DRC's science item and test development team will be responsible for overseeing the development of multiple-choice items and scenarios. Our team will deliver to PDE the required number of items, including scenarios for biology and chemistry, each year. Overseeing the process will be Mr. David Durette who has been the science development lead for Pennsylvania for the past four years. Mr. Durette will be supported by Mr. Robert Poppe, who has also been involved in the development of Pennsylvania Science assessments for many years. Mr. Poppe provided support to PDE during the development of the *Pennsylvania's Academic Standards*. He also provided support to PDE during PDE's initial development of test content blueprints and test designs for the PSSA science assessments, including developing, revising, and editing many of the science scenarios.

Reading (Literature)

The field test plan we propose is based upon a test design that is discussed under the subheading titled *Plan for Initial Field Testing* of our proposal. Upon award of the contract, DRC and WestEd development specialists would be pleased to meet with PDE to discuss test design and to reevaluate the number of passages provided each year.

Passage Development

DRC qualified reading staff will seek to ensure diversity in the selection of fictional and nonfictional reading passages that lend themselves to assessing various *Pennsylvania's Academic Standards*. DRC's designated GCA reading team will also seek to ensure that passages represent a variety of topics to include, but not limited to, the following: science, biography, technology, recipes, how-tos, and other informational topics; poetry, biography, and narratives for literary passages. For purposes of our cost proposal and in keeping with the philosophy of the GCA, DRC recommends developing 85% through 90% commissioned or public domain passages.

DRC uses a number of quality-control procedures to ensure the appropriateness and viability of passages. One step includes a review of the content accuracy and viability of the passages. *Appendix 2* contains our Passage Review Process and Fact Verification Sheet that presents the protocols through which passages are subjected and processed. We understand that PDE expects field tests and operational assessments that are balanced in terms of reading load, grade-level appropriateness, topic selection, and diversity. WestEd reading content, special education, and diversity specialists will support PDE through all stages of the process to create balanced reading assessments that are aligned to the *Pennsylvania's Academic Standards*.

In order to select passages for the literature assessment that meet the approved test and item specifications, we propose to use passage finders and writers who have previous experience with us, along with our reading test development specialists who have experience selecting and editing passages for large-scale reading assessments, including selecting passages for the GCA. DRC's designated GCA Reading Team Lead, reading test development specialists, and passage finders have had or currently have classroom teaching experience or have core content knowledge in the field of literature. The reading test development specialists, passage finders, and passage writers are also trained to use a variety of primary and secondary published sources, including magazines and books. Passage writers come from a pool of published authors, current teachers, and English majors. We believe the passage finders and writers are some of the best in the industry, and we believe PDE will be pleased with the passage selections we will provide.

Procedures for Passage Selection

Before beginning the process to secure passages, DRC proposes to meet with PDE. The purpose of the meeting will be to review our passage selection process. We propose that this meeting with PDE take place during the initial item development planning meeting to be held each year. At this meeting, we will confirm with PDE the number of passages per passage type, the passage criteria, and item specifications required by the program. As a result of our experience in selecting passages for Pennsylvania, we have learned that a meeting with PDE prior to beginning the passage selection process helps to ensure that expectations are clearly defined for us and are in alignment with the requirements of PDE. It is vital that our passage finders and writers and reading item and test development

specialists have a clear understanding of all special considerations required by the program. At this meeting, we will discuss with PDE our preliminary understanding of the types of passages required by the program. We will also present sample passages to PDE to secure feedback on our selections and to determine whether or not our initial sample selections are representative of the types of passages that PDE deems technically appropriate for each year's development.

Once we have met with PDE and have received PDE's approval to proceed with passage finding and passage writing, we will train the passage finders and writers. The passage training will include the following:

- Overview of the assessment program, including purpose of the program.
- General information concerning the number and types of passages needed.
- General and specific requirements, including the specifications of the GCA literature tests.
- Timelines for submission of passages to DRC.
- Specific information concerning how passages are to be written and prepared, including documentation of the source, type of passage, etc., as required by the criteria established by PDE.

In addition, during our passage training all passage finders and passage writers, including our own reading Test Development team members, will receive passage instructions and general guidelines. A sample of our general guidelines for passage selection and writing can be found below. The instructions will be customized to meet the needs of the GCA program.

General Guidelines for Passage Selection and Writing

Passages will have:

- The appropriate length for use and the appropriate density for high school level.
- Text that is rich enough to allow for the needed number of items to be generated.
- Text that will appeal to students at high school level.
- Appropriate subjects for high school level.
- Grade-appropriate vocabulary.
- Text structures that will be familiar for high school level.
- Text that is written in Standard American English. Fiction passages may contain colloquial expressions in dialogue, but these expressions should be appropriate.

Passage Readability

Evaluating the readability of a passage is essentially a judgmental process by individuals familiar with the classroom context and what is developmentally and linguistically appropriate for students at a high school level. Although readability indices will be computed and made available to PDE upon request, we believe that these indices measure different aspects of readability and can result in interpretive liabilities. Consequently, as we have historically practiced for the GCA, we recommend that the common readability formulas not be used in a rigid way, but that they be considered more informally to provide a factor for consideration in the selection of final passages for development. We recommend the use of experience-based judgments as outlined above.

Sources for Passages

The sources DRC's passage finders and reading item and test development specialists use or consider appropriate for use include newspapers; novels; trade books; anthologies of literature and poetry; short story collections; and young adult, and general magazines. The sources our passage finders typically avoid include Newbery Medal winning book titles, Caldecott Medal winning book titles, federal government forms, and selections from any basal reading series or from textbooks used within Pennsylvania. We also encourage our passage finders to avoid selecting passages that are older and contain outdated information, especially when collecting nonfictional materials and passages that are too popular or may have been used in reading/English curricula within Pennsylvania classrooms. In our initial meeting with PDE, we will discuss our preliminary plans for sources of passages and incorporate any suggestions and/or revisions PDE might have.

DRC will select any required literature (fictional and nonfictional) and graphics (e.g., drawings, timelines, photographs, graphs) based on the criteria established by PDE. We understand that literature passages, items, and graphics will also be reviewed for bias and sensitivity. In an effort to ensure that sufficient numbers of passages will survive PDE review, Content Review by committees of Pennsylvania educators, and the Bias, Sensitivity, Fairness Review by external experts, DRC's reading item and test development team will review each literature passage before it is submitted to PDE. DRC's reading item and test development team members will review each passage to determine the following:

- Match to Target Passage Types for the Literature GCA.
- Quality of the writing, including real-life authentic context that lends itself to high quality robust item development as required by the *Pennsylvania's Academic Standards*.
- Interest level and content appropriateness, including whether or not the content is meaningful and important for students.
- Accuracy of the information provided in the passage.

- Cultural diversity and freedom from issues of bias, fairness, and/or sensitivity.
- Grade-level appropriateness, including conceptual load, vocabulary, syntactic patterns, sentence length, and clarity.
- Reading level, as defined by what is typically found at the high school level.
- Use of illustrations, graphics, timelines, photographs, etc., including whether or not they are reproducible and adhere to the Principles of Universal Design.
- Other, as requested by PDE.

DRC acknowledges that all reading passages and graphics will be reviewed and approved by PDE prior to review by the external Content Review Committees and the external Bias, Sensitivity, Fairness Review Committee following a delivery schedule agreed upon by PDE and DRC. Our passage submissions will include the test-ready version, with any proposed art and graphics, as well as other information required by PDE.

Maintaining Passages Electronically

DRC will use our IDEAS item banking system to house all of the passages and eventually the items along with all accompanying illustrations, graphics, etc. IDEAS will be configured specifically for the GCA program. As passages are entered, information will be stored such as genre, word count, readability, course, etc. Other information can be held in our IDEAS item banking system for reading as requested by PDE. For further information about IDEAS, see the labeled *Item Banking System* of our proposal.

Permission Process

DRC understands that permissions, if necessary, must be obtained for all PDE-approved, previously published reading passages and graphics. DRC employs a full-time, highly experienced permissions editor who will oversee the permissions process. Our permissions editor, Mr. Ben Obler, has extensive experience seeking permissions for the PSSA, and he has secured permissions for the PSSA since 2003. He is detail-oriented, and he keeps accurate records throughout the process. He will seek to secure all necessary copyright permissions. Permissions for printed tests, computer administered tests, use in interpretive products, and electronic media such as CDs and release via the internet, will be obtained from the relevant parties. All rights will be secured on behalf of the Commonwealth of Pennsylvania and fees paid by DRC. Using databases and electronic tracking means, our permissions editor will seek permissions for developed passages and otherwise maintain copyright permissions for five years for all tests and ancillary products.

Our permissions acceptance rate is one of the best in the industry, since rarely is a passage denied usage. We are sure that PDE will be pleased with our permissions

process and rate of securing passages for the GCA. Historically, our permissions record has been extremely successful, and we will continue to provide this excellent permissions process and service to PDE.

Training Activities for Literature Item Writers

As in science, all levels of the DRC item and test development staff will participate in training before the development cycle begins. This training will be organized and conducted by Ms. Anne Kirpes, DRC literature and English composition content lead. The training will follow the same process as described in science, but it will also include special consideration of the passage genre and how the assessable standards relate to each genre. In addition, the set of items developed for a passage must span academic standards and depth-of-knowledge levels.

Training Literature Item Writers

As with science, DRC and WestEd use a combination of both in-house and contract content area item writers for GCA item writing. DRC and WestEd employ experienced writers who can develop technically sound and contextually rich items, and item writing assignments are based on the writer's content expertise. All contract content area item writers sign a statement in which they agree to treat all materials related to item development as confidential and not to disclose the content of the materials or the communication about item development.

Item writers receive training in the understanding of the Pennsylvania standards, the expectations for alignment to the Pennsylvania standards, the item specifications, the Pennsylvania style guide, depth-of-knowledge, the importance of avoiding cueing and overlap among items within a set of items for a given passage, and the use of IDEAS for item submission.

Writing Literature Items

As with mathematics and science, literature content area item writers will input items, stems, answer options, and distractor rationales directly into IDEAS, our item banking system. Other pertinent information associated with the item (e.g., depth of knowledge, estimated difficulty, focus) will also be entered.

Writing (English Composition)

DRC's English Language Arts item and test development team will be responsible for overseeing the writing prompt development and the development of the stimulus-based revising and editing passages, with multiple choice items. Prompts and items will be developed to meet Pennsylvania specifications. Each writing prompt and revising and editing multiple-choice item will have its associated metadata, as noted below. The trained content experts developing the writing prompts and the revising and editing items will ensure that prompts and items are grade-level appropriate. Prompts and items will be entered into IDEAS, our item banking system.

Writing Training Activities

Overseeing the process will be Ms. Cynthia Mann, who has been the writing content lead for the Pennsylvania Writing Assessment since 2004. DRC proposes to PDE that the prompts and the stimulus-based passages, with multiple-choice items, will be written by our content-area item and test development specialists, with the input and assistance of our performance assessment staff, Ms. Sue Drexler and Ms. Wendy Marik, as well as by contract prompt writers and stimulus-based passage writers from across the country who have content area knowledge and have written for us in the past.

Item writer training will follow the same process as described in mathematics, science, and literature, but it also includes special consideration for the development of prompts and the development of revising and editing stimulus-based passages, with multiple-choice items. For example, the training will focus upon important topics such as:

- Prompt-specific technical quality issues, such as writing prompts to be contextually relevant and developmentally appropriate in structure and content (e.g., text structure, topic),
- Bias and sensitivity issues in prompt construction and in the development of stimulus-based revising and editing passages with items,
- Understanding the Pennsylvania Mode-Specific Scoring Guidelines,
- Description of the prompt writing process and the stimulus-based passage development process, including how to write high-quality prompts, stimulus-based passages, and items to meet the requirements for special populations, and
- Guidelines for the development of writing prompts to ensure that an adequate number and type of prompt (informational and persuasive) is always available in the prompt bank.

Stimulus-Based Passages with Multiple-Choice Items

Each year DRC will develop the required number of stimulus-based passages, as required in the RFP. Each stimulus-based passage will have multiple-choice items associated with it. The items will measure the *Pennsylvania's Academic Standards 1.5.E* and *1.5.F* that measure quality of writing. Metadata elements include: associated stimulus-based passage, a unique item number, the content area, reporting category, standard, focus, Webb's depth of knowledge, estimated difficulty, and answer key.

All multiple-choice items will have four options with one and only one correct answer. Our development staff will not use "All of the above," "None of the above," or similar answer options. The correct response and all distractors will be from or based upon the stimulus-based passage.

Writing Prompts

Each year DRC will develop and deliver to PDE the required number of informational and persuasive writing prompts. The writing prompts will measure *Pennsylvania's Academic Standards* for Types of Writing (1.4.B informational and 1.4.C persuasive). Metadata elements include: a unique item number, type of prompt (mode), the content area, reporting category, and standard. Each prompt will also include its mode specific scoring guidelines and conventions guidelines.

All writing prompts will be developed and reviewed by DRC's English language arts team and scoring experts within DRC's Performance Assessment Center, specifically under the direction of Ms. Sue Drexler and Ms. Wendy Marik. Ms. Drexler and Ms. Marik have been involved in overseeing the scoring of the PSSA writing assessment for many years. They will also lend their experience for the scoring of the English Composition GCA. Both are knowledgeable of writing prompt development issues. They will lend their expertise to the development and review of writing prompts prior to delivery to PDE for review.

Social Studies (World History, U.S. History, and Civics and Government)

As with mathematics, science, reading, and writing, our social studies (World History, U.S. History, and Civics and Government) content area item writers will input items, stems, answer options, and distractor rationales directly into IDEAS, our item banking system. Other pertinent information associated with the item (e.g., depth of knowledge, estimated difficulty, focus) will also be entered.

Training Activities for Social Studies Item Writers

All levels of the WestEd item and test development staff will participate in training before the development cycle begins. This training will be organized and conducted by the social studies content lead. The training will follow the same process as described in science, reading, and writing, but it will also include special consideration of the development of social studies stimuli and item sets. It is critical that content area item writers and editors acknowledge and maintain the depth-of-knowledge balance throughout the item writing and editing process.

Training Social Studies Item Writers

Each content area writer will also be asked to pay careful attention to the readability of each item and check to ensure that the focus is upon the concepts, not upon reading. The goal is for each writer to write items and stimuli that are, to the greatest degree possible, independent of the assessment of reading; however, writers will receive instruction concerning grade-appropriate *social studies* vocabulary.

Writing Social Studies Items

Social Studies content area item writers will input items, stems, answer options, and stimuli directly into IDEAS, our item banking system. Other pertinent information associated with the item (e.g., depth of knowledge, estimated difficulty, focus) will also be entered.

WestEd's social studies item and test development team will be responsible for overseeing the development of multiple-choice items and stimuli. Their team will deliver to PDE the required number of items for World History, U.S. History, and Civics and Government each year.

Social Studies Stimulus

Stimuli will be developed for World History, U.S. History, and Civics and Government. Each is accompanied by a set of multiple-choice items. The social studies content will contain multiple types of stimuli, such as quality graphics, authentic data, and primary sources that are grade-level appropriate. These stimuli will be accessible to all students and model best practices in the classroom and will be developed to provide for a measurement of both process and content skills.

Security Procedures During Item Development

The maintenance of test security through the item development process is essential given the nature of high-stakes assessment. DRC and WestEd recognize the importance of security and take a series of steps to recognize and maintain security of test items through our physical plants, the maintenance of secure electronic environments, and secure file transfer.

At each item writing workshop (reading, mathematics, writing, science, and social studies), writers will be asked to sign a Confidentiality Letter specifying the confidentiality agreement and security regulations. The agreement will also outline the ownership regulations. DRC and WestEd understand that all work developed under this contract will be the sole property of PDE and the Commonwealth of Pennsylvania. DRC and WestEd also understand that no confidential materials related to the project will be released without PDE's explicit approval. All contract item writers sign a statement in which they agree to treat all materials related to item development as confidential and not to disclose the content of the materials or the communication about item development.

VII.B.1. English Composition Assessment Design

DRC has provided cost information as required by the RFP, in DRC’s Cost Submittal response document. The plans for an English Composition Assessment include 80% of the points based on open-ended items and 20% based on the selected response items. Our test designs for English Composition provide for 20 points per module with each of the 3 modules having an equal number of prompts and open-ended items.

Points Plan % OE & % MC	Unique Core MC per Form	Unique Core OE per Form	Total Core Items per Form		Total Core Points per Form
			MC	OE	
80/20	12	3 16 pt	12	3 16 pt	60

VII.B.2. Assessment Design Scenarios

DRC is pleased to present the following detailed designs for the operational assessments. The designs for all courses except for English Composition reflect the three ratio options requested in the RFP:

- a) 50% of points based on open-ended items and 50% of points based on multiple-choice items.
- b) 35% of points based on open-ended items and 65% of points based on multiple-choice items.
- c) 20% of points based on open-ended items and 80% of points based on multiple-choice items.

Algebra I, Algebra II, and Geometry

Our test designs outlined below for Algebra I, Algebra II, and Geometry provide for 20 points per module with each of the 3 modules having an equal number of Core, FT, and OE items.

Algebra I, Algebra II, and Geometry

Points Plan % MC & OE	Unique Core MC per Form	Core- to- Core Linking MC per Form	Embedded Field Test MC per Form	Unique Core OE per Form	Core- to-Core Linking OE per Form	Embedded Field Test OE per Form	Total Core Items per Form		Total Forms	Total Core Points per Form
							MC	OE		
50/50	21	9	15	3 4pt 3 2pt	3 4pt 0 2pt	3 4pt 3 2pt	30	6 4pt 3 2pt	5	60
65/35	24	15	15	2 4pt 1 3pt	1 4pt 2 3pt	3 4pt 3 3pt	39	3 4pt 3 3pt	5	60
80/20	33	15	15	2 4pt	1 4pt	3 4pt	48	3 4pt	5	60

Literature

Our test designs for Literature provide for 20 points per module with each of the 3 modules having an equal number of Core, FT, and OE items.

Points Plan % MC & OE	# of Unique Core Passages per Form	Unique Core MC per Form	Unique Core OE per Form	# of Core-to-Core Passages per Form	Core-to-Core Linking MC per Form	Core-to-Core Linking OE per Form	# of Field Test Passages per Form
50/50	6	21	3 4pt 3 2pt	3	9	3 4pt 0 2pt	3
65/35	6	24	2 4pt 1 3pt	3	15	1 4pt 2 3pt	3
80/20	6	33	2 4pt	3	15	1 4pt	3

Points Plan % MC & OE	Embedded Field Test MC per Form	Embedded Field Test OE per Form	Total Core Items per Form			Total Forms	Total Core Points per Form
			Passages	MC	OE		
50/50	15	3 4pt 3 2pt	9	30	6 4pt 3 2pt	5	60
65/35	15	3 4pt 3 3pt	9	39	3 4pt 3 3pt	5	60
80/20	15	3 4pt	9	48	3 4pt	5	60

Biology and Chemistry

Our test designs for Biology and Chemistry provide for 20 points per module with each of the 3 modules having an equal number of Core, FT, and OE items.

Points Plan % MC & OE	# of Unique Core Scenarios per Form	Unique Core MC per Form	Unique Core OE per Form	# of Core-to-Core Scenarios per Form	Core-to-Core Linking MC per Form	Core-to-Core Linking OE per Form
50/50	3	21	3 4pt 3 2pt	0	9	3 4pt 0 2pt
65/35	3	24	2 4pt 1 3pt	0	15	1 4pt 2 3pt
80/20	3	33	2 4pt	0	15	1 4pt

Points Plan % MC & OE	# of Field Test Scenarios per Form	Embedded Field Test MC per Form	Embedded Field Test OE per Form	Total Core Items per Form			Total Forms	Total Core Points per Form
				Scenarios	MC	OE		
50/50	1	15	3 4pt 3 2pt	3	30	6 4pt 3 2pt	5	60
65/35	1	15	3 4pt 3 3pt	3	39	3 4pt 3 3pt	5	60
80/20	1	15	3 4pt	3	48	3 4pt	5	60

World History, U.S. History, and Civics and Government

Our test designs for World History, US History and Civics & Government provide for 20 points per module with each of the 3 modules having an equal number of Core, FT, and OE items.

Points Plan % MC & OE	# of Unique Core SS Item Sets per Form	Unique Core MC per Form	Unique Core OE per Form	# of Core-to-Core SS Item Sets per Form	Core-to-Core Linking MC per Form	Core-to-Core Linking OE per Form
50/50	3	21	3 4pt 3 2pt	0	9	3 4pt 0 2pt
65/35	3	24	2 4pt 1 3pt	0	15	1 4pt 2 3pt
80/20	3	33	2 4pt	0	15	1 4pt

Points Plan % MC & OE	# of Field Test SS Item Sets per Form	Embedded Field Test MC per Form	Embedded Field Test OE per Form	Total Core Items per Form			Total Forms	Total Core Points per Form
				SS Item Sets	MC	OE		
50/50	1	15	3 4pt 3 2pt	3	30	6 4pt 3 2pt	5	60
65/35	1	15	3 4pt 3 3pt	3	39	3 4pt 3 3pt	5	60
80/20	1	15	3 4pt	3	48	3 4pt	5	60

VII.B.3. Items for Assessments

DRC and WestEd will provide all items appropriate for all the GCA assessments. When appropriate, the assessments will contain passages, graphics, and scenarios. As outlined in this section, items will be written to Dr. Norman Webb’s depth-of-knowledge levels 2 or 3. DRC and WestEd acknowledge that the items must be secure within Pennsylvania and may be released with only written permission by

PDE. As outlined earlier in our proposal in the Item Development section, *Subheading VII.B*, all items will be reviewed for content alignment, grade-level appropriateness, difficulty, depth-of-knowledge, bias, sensitivity and fairness.

Prepare Items for Review by PDE

All developed items, with associated passages, stimulus prompts, scenarios, graphics, and all item characteristics such as item code, estimated difficulty, depth-of-knowledge level, distractor analysis, focus, scoring guidelines, etc., will be prepared for review by PDE. We understand the importance of the review, and we will work with PDE to establish schedules and procedures that will facilitate PDE's review of all items, prompts, scenarios, etc., prior to reviews by committees of Pennsylvania educators. For the review, PDE will have direct access to IDEAS, our item banking system, and PDE may elect to review items electronically in "real time" using IDEAS. Our item-development team will work with PDE to meet PDE's review process requests. Please see below for a discussion of IDEAS.

Item Banking System

DRC is proud to have developed one of the most comprehensive item development, banking, and form construction software applications in the industry today. Our item banking system, IDEAS (Item Development and Educational Assessment System, patent pending), was designed and built to provide **a single, consistent repository to house all information relating to test items, passages, and test forms**. This approach allows staff from Test Development, Psychometrics, Document/Graphics Design Group, and Printing Services to work seamlessly together through a common, user-friendly system.

IDEAS provides all the functionality required to take an item from authoring to review to forms construction and publications/printing. **The flexibility of the data design allows for complete client customization of the data elements captured and associated with the items**. These data elements can range from simple item characteristics such as content area to the most complex item statistics, rubrics, and difficulties.

Some of the key features and functionality of IDEAS are highlighted below.

IDEAS Key Features

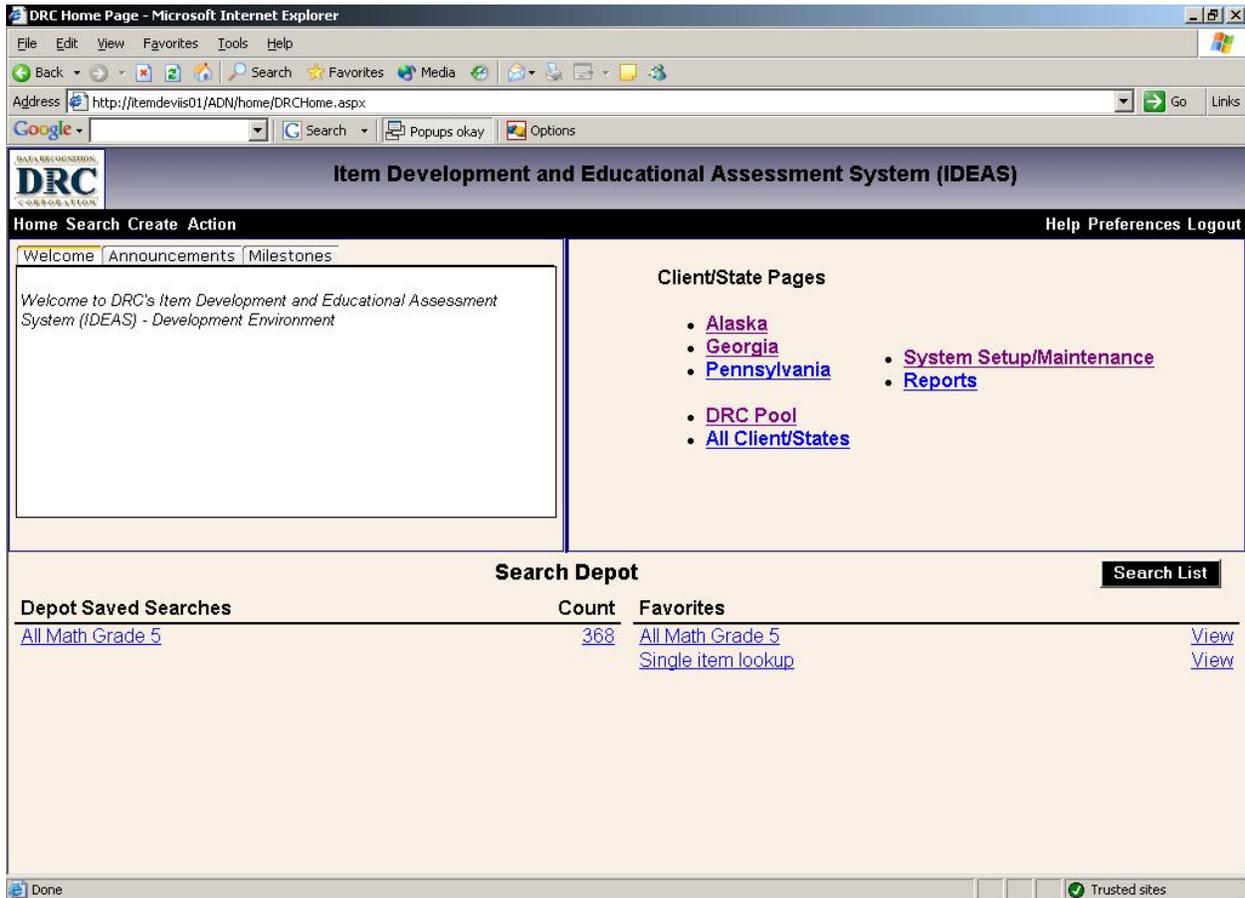
- **Highly secure**—Protected by DRC’s corporate-wide information technology security controls. Password assignment, sign-in process authenticates users based on project roles. Authorized users can access only the areas pertinent to their roles. An audit trail feature displays and documents user changes.
- **Web-based**—Allows authorized users to access its capabilities from a variety of locations, including PDE offices and committee review locations.
- **Fully searchable database**—All item and passage data can be easily managed, searched for, and retrieved. Authorized users have a range of search capabilities, including the ability to define and save their own custom searches.
- **Tracking individual items and passages**—Unique item and passage identifiers enable tracking throughout the item development and forms development processes.
- **Tracking development progress**—The development of the item bank can be easily and accurately monitored to ensure content coverage across and within standards.
- **Customizable real-time progress reports**—IDEAS produces a wide variety of customizable reports to meet the needs of the GCA.
- **Documentation of item and passage history**—Updates to items and passages occurring at any stage, including external committee reviews, are recorded. An historical record of all changes/revisions is kept throughout the life of the item/passage.
- **Streamlined test form development and publishing**—Using approved and selected items and passages, the system automatically generates print-ready files. This single repository system allows DRC staff from all areas to seamlessly work

IDEAS provides clear benefits to our assessment clients.

Assessment Program Benefits of IDEAS

- **Remote accessibility**—Web-based system will allow authorized PDE staff to access IDEAS from department offices on a 24/7 basis. PDE will be able to search and view items and passages and print item/passage cards. Items and passages flagged by DRC will indicate the need for PDE review and response. PDE will be able to input questions or revisions.
- **Enhanced item/passage accuracy**—A historical record of all changes and revisions to the items will ensure that items and passages used on tests accurately reflect the input of PDE and review committees
- **Real-time development progress reports**—PDE and DRC will have the ability to monitor the development of the item bank to ensure development blueprints and specifications are met. Fully customized reports will reflect the real-time status of the item bank at any given time.
- **Enhanced test-form accuracy and efficiency**—The single repository system streamlines the test form development, publishing, and printing processes. Automatically generated print-ready files ensures item and passage accuracy.

The figure below illustrates the system's home page.



IDEAS Home Page Screen

System Accessibility

The system is web-based to allow appropriate users to access its capabilities from a variety of locations. Security of the system and residing data is of the utmost importance. Appropriate personnel are authenticated via unique log-ins and passwords, as well as secure socket layer (SSL) technology to maintain its integrity. This authentication also becomes the basis for determining appropriate user-authority levels in the system. Persons allowed into the system are limited to only those functions that are necessary to perform their jobs.

In order for authorized users to easily manage and retrieve the immense amount of data contained in IDEAS, the system provides users with a range of search capabilities. A number of pre-defined searches and reports allow for rapid access to the most commonly used information. Additionally, system users have the capability to define and save their own custom searches based on nearly any field contained in the database whether it relates to items, passages, or forms. This provides an unlimited set of possibilities to users who require the ability to look at data from any angle.

PDE staff will have secure, remote, 24/7 access to IDEAS in order to search and view items and passages, as well as print item cards. The system can also be used to facilitate discussion between DRC and PDE regarding items and passages. DRC staff members will be able to attach flags to items and passages; these flags will be readily viewable and searchable by PDE staff and will indicate the need for PDE review and response. PDE staff will have the capability to insert questions or proposed revisions as comments and associate them with individual items and/or passages. When items or passages need revision, PDE will be able to print the specific item cards, make edits, and send to the appropriate DRC team members. This method will provide for a collaborative and efficient item/passage review and revision process.

IDEAS Security

Security for IDEAS will be provided through a Secure Socket Layer (SSL) protocol for information transfer over the Internet and Windows authentication for user access to the system. The system will be protected by a password assignment and sign-in process that authenticates users based on each person's role on the project. The item bank will have an audit trail feature that displays user changes to items and documents these changes. Authorized users will have access only to portions of the electronic item bank pertinent to their roles. For example, mathematics test development specialists will not be allowed to view science items, while others may be restricted to read-only access. Electronic item and form information stored in IDEAS will remain secure until written authorization has been received from the appropriate PDE contact to securely delete all such information.

IDEAS for Item Development

IDEAS will be used for all item and passage authoring for the GCA. After an item is submitted, the appropriate style is applied to the item. The system then allows for editing of the item by appropriate DRC and PDE personnel. The system maintains the item exactly as it will be presented on the test form.

DRC's test development team will ask each content area item writer and passage writer/finder to document specific information. This information will be provided on an item-writer/passage-writer template and will be entered electronically into our IDEAS item bank, where each item and passage is assigned a unique identifier. Item-level/passage-level associations will establish links as necessary to associated artwork, items/passages, and related items. These identifiers will allow IDEAS to be used to track items electronically and securely throughout the item development process and subsequent forms development process. In addition, through the use of IDEAS, DRC Test Development team members, and PDE staff members, if desired, will be able to track the development progress of items and passages. DRC test development specialists will monitor the development of the item bank to ensure content coverage across and within standards to provide breadth and depth of content.

Tracking reports can be generated and provided to PDE, upon request. The tracking report will show the current state of item and passage development, including number of items and distribution of items across standards. It will also show the complete item pool at any given point in the development process. IDEAS will ensure that updates to items occurring at any stage of the process, including external committee reviews, are recorded. Changes are noted and a historical record of all changes/revisions to the item is kept throughout the life of the item. The historical record will be available to PDE at any time.

The next figure illustrates an item authoring screen displaying a non-secure item owned by DRC.

The screenshot shows a web browser window titled "Item Maintenance - Microsoft Internet Explorer" with the address bar displaying "http://itemdeviis01/ADN/Item/ItemEdit.aspx?itemid=501205". The main content area is titled "Item Development and Educational Assessment System (IDEAS) - DRC Pool". The interface includes a navigation menu with "Home Search Create Action" and "Help Preferences Logout". Below this, there are tabs for "Item Details", "Passage Details", "Rubrics", "Graphics", "Attachments", "Notes", "Exemplar Responses", and "Statistics". The "Item Details" tab is active, showing fields for "Item ID" (501205), "Item Status" (Author), "Content Area" (SCIENCE), "Grade" (11), and "Save #" (11). The "Item Characteristics" section on the left includes dropdown menus for "Item Type" (Multiple Choice), "Subject" (Chemistry), "Key" (C), "Points" (1), "Est. Difficulty Level" (Medium), "Cognitive Level", "Bloom's Level" (2), "Depth of Knowledge" (2), "Calculator" (Neutral), "Category 1", "Category 2", and "State-Specific" (False). The main content area displays a question labeled "1." with two diagrams of atomic structures. The first diagram is labeled "sodium" and shows a nucleus with "P: 11" and three concentric electron shells. The second diagram is labeled "chlorine" and shows a nucleus with "P: 17" and three concentric electron shells. Below the diagrams is the question text: "Which statement describes what happens to the electrons in sodium and chlorine when table salt (sodium chloride) is formed?" and four multiple-choice options: A. Sodium and chlorine electrons are shared. B. Sodium and chlorine electrons fuse together. C. Sodium loses an electron, and chlorine gains an electron. D. Sodium electrons gain energy, and chlorine electrons lose energy. At the bottom of the screen are buttons for "Print", "Preview", "Open Editor", "Save", and "Cancel".

IDEAS Item Authoring Screen

All items will be written and field-tested specifically for Pennsylvania and will be the property of PDE. DRC and WestEd will develop items in sufficient quantities per the Assessment specifications in Appendix A as described below.

Mathematics (Algebra I, Algebra II, and Geometry)

Mathematics Multiple-Choice Items

All mathematics multiple-choice items will have four options with one and only one correct answer. WestEd item development staff will not use “All of the above,” “None of the above,” or answer options similar to these constructions. For each item, our writers will provide a rationale for all distractors. Distractors will be plausible choices that represent common errors and misconceptions in student reasoning.

Mathematics Open-Ended Items

The mathematics open-ended items will be scored on either a 0–4 scale, a 0–3 scale, or a 0–2 scale, depending on the ratio design chosen by PDE. The mathematics open-ended items will be written to present the examinees with an opportunity to demonstrate their strategic thinking.

Each open-ended item will include item-specific scoring guidelines. The scoring guidelines are initially written by item writers who have been specifically trained to develop guidelines. Subsequently, all open-ended items, with scoring guidelines, are also meticulously reviewed and further developed by WestEd’s mathematics content team and DRC scoring experts within our Performance Assessment Centers, specifically under the direction of Mr. Warren Hite. Mr. Hite has been involved in overseeing the scoring of mathematics items for over ten years. He is very knowledgeable of development issues related to open-ended items, and he will lend his expertise throughout the open-ended item development process.

Science (Biology and Chemistry)

Science Multiple-Choice Items

DRC will develop multiple-choice science items for biology and chemistry, and all developed items will meet Pennsylvania specifications. Each science item will have its associated metadata, including: a unique item number, the content area, reporting category, academic standard, sub-anchor, eligible content, item type, Webb’s depth of knowledge, estimated difficulty, and answer key. The trained content experts developing the items will ensure that items are grade-level appropriate.

All science multiple-choice items will have four options with one and only one correct answer. Our development staff will not use “All of the above,” “None of the above,” or answer options similar to these constructions. For each item, our writers will provide a rationale for all distractors.

Science Open-Ended Items

The science open-ended items will be scored on either a 0–4 scale, a 0–3 scale, or a 0–2 scale, depending on the ratio design chosen by PDE. Open-ended items also will include all relevant metadata, including: a unique item number, the content area, reporting category, academic standard, sub-anchor, eligible content, item type, points possible, Webb’s depth of knowledge, estimated difficulty, and associated scenario. Item-specific scoring guidelines will be provided for each open-ended item.

Each open-ended item will include item-specific scoring guidelines. The scoring guidelines are initially written by item writers who have been specifically trained to develop guidelines. Subsequently, all open-ended items, with scoring guidelines, are also meticulously reviewed by DRC’s science content team and scoring experts within DRC’s Performance Assessment Center, specifically under the direction of Ms. Violeta Lee. Ms. Lee has more than ten years of experience in science assessment design, development, evaluation, and scoring, and more than eight years of direct Pennsylvania assessment handscoring experience.

Science Scenarios

Scenarios will be developed for biology and chemistry. Each scenario is accompanied by a set of multiple-choice items. In addition, each science scenario will contain a real-world topic or common theme that is assessed by a variety of items with diverse levels of difficulty and cognitive demands. DRC has proven experience in crafting science scenarios for the Pennsylvania science assessments to assess core science concepts aligned to the eligible content. The science scenarios will contain multiple types of displays (stimuli) such as quality graphics, authentic data, and experimental designs that are grade-level appropriate. These displays will be accessible to all students and model best practices in the classroom. In addition, the science scenarios will fully address the interdisciplinary nature of science and science content standards that will create stronger connections of core science concepts. The scenarios will be developed to provide for a measurement of both process and content skills.

Literature

Literature Multiple-Choice Items

DRC will secure all passages and provide passages to WestEd for the development of items for the literature assessment. All developed items will meet Pennsylvania reading specifications. For purposes of our cost proposal, each passage will have multiple-choice items and open-ended items.

WestEd will develop multiple-choice items providing metadata for each item. Metadata elements include: associated passage with type, word count, readability, etc.; a unique item number; the content area; reporting category; academic standard; sub-anchor; eligible content; focus; item type; Webb’s depth of knowledge; estimated difficulty; and answer key.

All multiple-choice items will have four options with one and only one correct answer. Our development staff will not use “All of the above,” “None of the above,” or similar answer options. The correct response and all distractors will be from/or based upon information in the passage, except in the case of vocabulary items, for which answer options need not come from the passage.

Literature Open-Ended Items

WestEd will develop open-ended items to be scored on either a 0–4 scale, a 0–3 scale, or a 0–2 scale, depending on the ratio design chosen by PDE. Open-ended items will include all of the same metadata as provided for multiple-choice items except for answer key.

The open-ended items provide for the assessment of students’ reading comprehension in ways that multiple-choice items cannot. As part of their response, students are required to supply relevant supporting examples or information from the text. It is critical when developing these items that focus be on appropriate passage-based inferences and conclusions rather than on reflections or connections that presume too much about students’ common experiences.

Each open-ended item will include item-specific scoring guidelines. The scoring guidelines are initially written by item writers who have been specifically trained to develop guidelines. Subsequently, all open-ended items, with scoring guidelines, are also meticulously reviewed and further developed by DRC’s reading content team and scoring experts within DRC’s Performance Assessment Center, specifically under the direction of Ms. Sue Drexler and Mr. Nicholas Hook. Ms. Drexler and Mr. Hook have been involved in overseeing the scoring of PSSA reading items for over ten years and will lend their experience for the Literature GCA. They are both very knowledgeable of development issues related to open-ended items.

Social Studies

Social Studies Multiple-Choice Items

WestEd will develop multiple-choice social studies items and items sets (comprised of 4 – 6 items; 1 item set per module) for World History, U.S. History, and Civics and Government; all developed items will meet Pennsylvania specifications. Each social studies item will have its associated metadata, including: a unique item number, the content area, reporting category, academic standard, sub-anchor, eligible content, item type, Webb’s depth of knowledge, estimated difficulty, and answer key. The trained content experts developing the items will ensure that items are grade-level appropriate.

All social studies multiple-choice items will have four options with one and only one correct answer. Our development staff will not use “All of the above,” “None of the above,” or answer options similar to these constructions. For each item, our writers will provide a rationale for all distractors.

Social Studies Open-Ended Items

The social studies open-ended items will be scored on either a 0–4 scale, a 0–3 scale, or a 0–2 scale, depending on the ratio design chosen by PDE. Open-ended items also will include all relevant metadata, including: a unique item number, the content area, reporting category, academic standard, sub-anchor, eligible content, item type, points possible, Webb’s depth of knowledge, estimated difficulty, and associated scenario. Item-specific scoring guidelines will be provided for each open-ended item.

Social studies items will follow the similar item-specific scoring guidelines as other contents. Subsequently, all open-ended items, with scoring guidelines, are also meticulously reviewed by WestEd’s social studies content team and scoring experts within DRC’s Performance Assessment Center, specifically under the direction of Mr. Bob Dzandzara, World History, United States History, and Civics and Government Handscoring Lead. Mr. Dzandzara has 19 years of experience working on large-scale assessment projects.

VII.B.4. Quality of Items

GCA items, prompts, scenarios, passages, etc., will be reviewed by content-area item and test development specialists and content-editorial specialists for technical quality; match to anchor/standard; bias, sensitivity, fairness; depth of knowledge; estimated difficulty; adherence to the Principles of Universal Design; etc. Two additional editors will provide an independent review. The aim for this team approach is to conduct a multi-tiered internal review of all items, prompts, scenarios, etc., prior to submission for review by PDE and then, with approval by PDE, prior to submission for external committees to ensure that all items align with the *Pennsylvania Academic Standards* for Reading, Mathematics, Writing, Science, and Social Studies and adhere to PDE’s requirement for high-quality items.

DRC and WestEd understand the expectation of PDE to receive items and/or scoring guides that require little or no edits and are aligned to the Pennsylvania GCA Anchors. In addition, a chart will be submitted one month after the start of the contract that will identify expected item development by assessment anchors for the first three GCA courses. In order to meet this expectation DRC’s quality process is outlined below.

Our content item development and editorial team, including two additional independent editors, will review all items, prompts, scenarios, etc., to ensure that they possess the following characteristics:

- Content alignment or congruence with the knowledge and skills specified in the anchors or standards.
- A range of estimated difficulty levels.
- Appropriate grade-level vocabulary, subject matter, and assumed student knowledge.

- Freedom from issues or concerns for bias, sensitivity, or fairness.
- Accessibility, following the Principles of Universal Design.
- Correct grammar, usage, and structure/format.

Review and Revise Items Prior to Submission to PDE (Internal Editing Checks)

As a part of our internal review of the items, DRC and WestEd’s item and test development team members and graphic specialists ensure that item art can be reproduced clearly and accurately when test booklets are printed or electronically displayed.

Test specifications will be reviewed to identify any potential display requirements that may present challenges in a print or electronic display environment. Display tolerances are impacted by line thickness, percent screening for shading, specialized fonts and symbols, photographs, color, and paper type. These are defined in the early stages of the item and test development process to help guide the delineation of style requirements and specifications.

Item art is produced using vector graphics that allow for scalar adjustments without the breakdown of image clarity that is common with lower quality bit-mapped formats. DRC’s multi-tiered quality assurance process makes certain converted item art is carefully compared to the original format throughout the item and test development and production process.

Displaying High-Quality Art

The display of high-quality art in tests does not end with art production and the application of Universal Design principles. The medium for display and the conversion or transformation of the artist’s work to this medium are also given careful consideration.

Employing well-trained, highly competent professionals is a key reason for our success. However, our commitment to high-quality, error-free items requires that our procedures and processes are state-of-the-art and as reliable and thorough as the professionals who implement them. As stated previously in the proposal, to begin the process, DRC will first collaborate with PDE on the style guide **to ensure consistent application of preferences and expectations**. The current PSSA Style Guide will be used to develop the GCA Style Guide. This style guide will identify not only the font and font size to be used on all forms, but also the graphic composition, style, and composition of all items and stimuli. The use of a Pennsylvania-specific style guide will ensure that detailed specifications for item and materials development are available to all staff members that collaborate on item and materials development for the Pennsylvania GCAs. It will also serve as the principal resource document to facilitate item and testing materials discussion between the PDE and DRC.

All items and test materials developed at DRC undergo stringent proofing quality assurance procedures. Before any item or test material is submitted to PDE, DRC's editing team, led by **Ms. Elizabeth Joyce, Lead Editor**, will be responsible for coordinating word-for-word proofreading; at least two editors will perform two independent word-for-word reviews of materials. The editing team will conduct a final "three-tier" proofreading of all test booklets, answer documents, and manuals to confirm that all directions in each form work in concert and are accurate and easy to follow.

DRC will follow the same quality control procedures as we do for test material development. In other words, after each step of the committee process, DRC's test development editing team will be responsible for coordinating word-for-word proofreading; at least two editors will perform two independent word-for-word reviews of items to ensure that all requested revisions to items have been made. If the changes require proofing against answer documents or manuals, the editing team will also conduct a "three-tier" proofreading of the test items and prompts, answer documents, and manuals to confirm that all work in concert and are accurate.

Before PDE reviews any item, the item has been through multiple levels of review, not counting the originating author. After an author submits an item in IDEAS, the DRC-developed item banking system, an intake editor decides whether the item may move forward to first-level editing. (The intake process may also result in the item being rejected or sent back to the author for revision.) Two subsequent rounds of content editing follow the first review. The content director then gives the item another review before it is submitted to PDE for review. PDE staff will review all items and provide DRC and WestEd with revisions prior to the content review meetings. All revisions will be made per PDE request.

These well-developed procedures are optimized via an item management system that facilitates the authoring, processing, routing, tracking, and formatting of test items. We propose to carry out the development work for this contract using IDEAS, the custom-designed item management system of DRC. PDE may wish to review the items in "real time" using our IDEAS item banking system throughout any stage of the process. IDEAS will capture and track all edits requested by PDE and/or subsequent revisions to items, per PDE request, by DRC and WestEd. Further information concerning IDEAS, our item banking system, can be found later in this section under the heading titled *Item Banking System*.

Transition of GCA Item Bank

DRC will work with PDE to ensure the continuity of the assessment system beyond the end of the contract period. All item, administration, and statistical data will be converted into an appropriate, PDE-approved format from DRC's electronic item bank. Our Test Development, Psychometric, and Information Services staff will work together with the PDE to ensure accuracy in the transition. Our electronic item bank solution stores items in formats that are easily

translatable to other publishing packages or capable of being easily exported. We understand the contents of the GCA item bank are the property of the PDE and must be transferable to the PDE. All transferred data and information will be transferred via CD-ROM/DVD or a secure file transfer protocol (SFTP) site, utilizing all necessary security measures, including encryption. In addition, hardcopy of items in the item bank will also be distributed.

VII.B.5. Quantity of Items

DRC and WestEd will develop and provide items in sufficient quantity to cover six years of assessments as outlined in the tables below. The tables represent the number of items that must be approved for use on a field test. DRC’s item development plan takes into account item attrition resulting from external reviews (e.g., content reviews).

Algebra I, Algebra II, and Geometry Standalone Field Test Totals

Points Plan % MC & OE	Total # of Unique Field Test OE Items		Total # of Unique Field Test MC Items	Total Per Course Year 1
	54 36	4pt 2pt		
50/50	54 36	4pt 2pt	195	90 OE 195 MC
65/35	36 36	4pt 3pt	249	72 OE 249 MC
80/20	36	4pt	303	36 OE 303 MC

Algebra I, Algebra II, and Geometry Embedded Field Test Totals

Points Plan % MC & OE	Field Test Forms	Annual Embedded Field Test MC Items	Annual Embedded Field Test OE Items	
			27 27	4pt 2pt
50/50	9	135	27 27	4pt 2pt
65/35	11	165	33 33	4pt 3pt
80/20	15	225	45	4pt



**US History, World History, and Civics & Government
Standalone Field Test Totals**

Points Plan % MC & OE	Total # of Unique SS Item Sets	Total # of Unique Field Test OE Items		Total # of Unique Field Test MC Items	SS Item Set Description	Total Per Course Year 1
50/50	32	72 48	4pt 2pt	240	2 MC and 1 OE per set	240 MC 120 OE
65/35	32	48 48	4pt 3pt	312	2 MC and 1 OE per set	312 MC 96 OE
80/20	32	48	4pt	384	3 MC and 1 OE per set	384 MC 48 OE

**US History, World History, and Civics & Government
Embedded Field Test Plan**

SS Item Sets	Field Test Forms	Annual Embedded Field Test MC Items	Annual Embedded Field Test OE Items	
9	9	135	27 27	4pt 2pt
11	11	165	33 33	4pt 3pt
15	15	225	45	4pt

Literature Standalone Field Test Totals

Points Plan % MC & OE	Total # of Unique Passages	Total # of Unique Field Test OE Items		Total # of Unique Field Test MC Items	Total Per Course Year 1
50/50	75	72 48	4pt 2pt	240	240 MC 120 OE
65/35	75	48 48	4pt 3pt	312	312 MC 96 OE
80/20	75	48	4pt	384	384 MC 48 OE

Literature Embedded Field Test Totals

Points Plan % MC & OE	Passages	Field Test Forms	Annual Embedded Field Test MC Items	Annual Embedded Field Test OE Items	
50/50	27	9	135	27 27	4pt 2pt
65/35	33	11	165	33 33	4pt 3pt
80/20	45	15	225	45	4pt

Biology and Chemistry Standalone Field Test Totals

Points Plan % MC & OE	Total # of Unique Scenarios	Total # of Unique Field Test OE Items		Total # of Unique Field Test MC Items	Total Per Course Year 1
50/50	32	72 48	4pt 2pt	240	240 MC 120 OE
65/35	32	48 48	4pt 3pt	312	312 MC 96 OE
80/20	32	48	4pt	384	384 MC 48 OE

Biology and Chemistry Standalone Field Test Totals

Points Plan % MC & OE	Total # of Unique Scenarios	Total # of Unique Field Test OE Items		Total # of Unique Field Test MC Items	Total Per Course Year 1
50/50	32	72 48	4pt 2pt	240	240 MC 120 OE
65/35	32	48 48	4pt 3pt	312	312 MC 96 OE
80/20	32	48	4pt	384	384 MC 48 OE



English Composition Standalone Field Test Totals

Total # of Unique Field Test OE Items		Total # of Unique Field Test MC Items
90	16pt	Approximately 264

Item Totals by Content for Entire Contract

	Algebra I	Algebra II	Geometry	Biology	Chemistry	English Comp.	Literature	US History	World History	Civics & Government
50/50	870 MC 414 OE	870 MC 414 OE	870 MC 414 OE	780 MC 336 OE	645 MC 282 OE	N/A	645 MC 282 OE	780 MC 336 OE	780 MC 336 OE	645 MC 282 OE
65/35	1074 MC 402 OE	1074 MC 402 OE	1074 MC 402 OE	972 MC 360 OE	807 MC 294 OE	N/A	807 MC 294 OE	972 MC 360 OE	972 MC 360 OE	807 MC 294 OE
80/20	1428 MC 261 OE	1428 MC 261 OE	1428 MC 261 OE	1284 MC 228 OE	1059 MC 183 OE	264 MC 90 OE	1059 MC 183 OE	1284 MC 228 OE	1284 MC 228 OE	1059 MC 183 OE

VII.B.6. Maximum Use of Items

The designs for maximum use of items are provided in *Subheading VII.B.2. Assessment Design Scenarios*. Items may be repeated, but not in two consecutive years. DRC understands that PDE intends to release a portion (10%) of items each year.

VII.B.7. Review Committees

Prepare all Materials for Content Item Reviews

It is our belief that the educators of Pennsylvania offer invaluable insight into the passages and items appropriate for their students, and we anticipate continued success with providing support to PDE in this process. After all proposed items, prompts, scenarios, etc., have been reviewed, revised per PDE’s request, and subsequently accepted by PDE, they will be prepared for presentation to review committees comprised of Pennsylvania educators. Only passages and items that are approved by PDE will be included in the final pool for review by the external committees of Pennsylvania educators.

For each course GCA committee, items, passages, scenarios, and prompts will be secured in binders and will typically be printed one per page and will include all information such as, but not limited to, what each item is measuring (Academic Standard and eligible content for mathematics, reading, science, and social studies), focus, depth-of-knowledge level, answer key or scoring guideline, distractor rationale, and other information as requested by PDE. All passages and stimulus-based prompts will also be included in the binders.

Support PDE with the Content Item Review Committee Meeting Process

Committees of Pennsylvania educators will review all items (e.g., content-area appropriateness, curricular alignment of the items, cognitive demand and rigor alignment, bias and sensitivity, adherence to the Principles of Universal Design). In addition, scoring guidelines for open-ended items and writing prompts will also be reviewed. Our content-area specialists will support PDE as facilitators and note takers.

DRC understands that the Content Item Review meetings will take place in Pennsylvania and will consist of Pennsylvania educators and other content matter experts as specified in the RFP. The table below shows the number of committee members we propose based on RFP requirements. The table also shows the number of days for each Content Item Review Committee meeting. Please refer to *Appendix 12* for more detailed information about these committees, including the years in which they will be convening.

Content Area	GCA Course	Number of Committee Members	Number of Days for the Meeting
Mathematics	Algebra I	15	5
	Algebra II		
	Geometry		
Science	Biology	15	3
	Chemistry	15	3
English/Language Arts	English Composition	15	5
	Literature	15	4
Social Studies	US/World History	15	4
	Civics & Government	15	4

DRC understands that PDE currently has a file of possible content reviewers, that PDE will establish all committees, and that the diversity of participants at each meeting will match the specifications for committee composition as detailed in the RFP. DRC understands that PDE has a committee participant database and will assist PDE in selecting the committee members should PDE request. For example, if PDE requests, we could work with PDE to draft an invitation to potential review committee members, including an overview of the task, time, and date of meetings, and all other relevant information. In addition, the invitation might include a questionnaire designed to capture information about each potential review committee member, including number of years of classroom teaching experience, educational background, item content review experience, diversity, gender, and geographical location within the Commonwealth of Pennsylvania. We would carefully proof the invitation and questionnaire and provide both to PDE for review and final approval.

DRC understands the challenges involved with recruiting meeting participants. DRC would like to discuss with PDE, upon contract award, the use of incentives (beyond ACT 48 credits) to bolster participation.

DRC's proposal includes the following additional activities associated with supporting PDE with the content item review committee process:

- Coordinating and establishing all meeting logistics, including hotel procedures, meeting rooms, computers, copier capability, etc.
- Corresponding with committee members, including travel arrangements, meeting announcements, etc., should PDE request. (Note: All written communication will be reviewed and approved by PDE prior to being sent to content review committee members.)
- Paying expenses for committee members and compiling the information necessary for professional development hours.
- Obtaining all supplies, including paper, pencils, flip charts, name tags, travel reimbursement forms, security/confidentiality documents, and review tracking documents and approval/revision tally sheets, and other documents, as requested by PDE.
- Supporting PDE staff with the development of the Content Item Review training materials, including PowerPoint training presentation slides, should PDE request such support. The information in these materials will include training participants to review items for content alignment, rigor level alignment, technical design, issues, and Principles of Universal Design.
- Supporting PDE staff with the facilitation of the committee review, including compiling the feedback information.

Costs for all relevant committee review meeting tasks and expenses as required by the RFP are included in our *Cost Submittal*, provided under separate cover.

Proposed Staff for the Content Item Review Committee Meetings

DRC understands that the Content Item Review Committee meetings will be facilitated by PDE, DRC, and WestEd. As a result, DRC and WestEd propose to provide the necessary number of content-area experienced staff to support the completion of all review tasks in the time available. Staff will include content-area item and test development specialists and editors. One additional staff member from DRC will serve as the meeting logistics coordinator. We understand the importance of providing the Commonwealth of Pennsylvania with staff members who have had previous experience facilitating content item review meetings and/or supporting the facilitation of content item review meetings in Pennsylvania. We believe that PDE will be pleased with the expertise of our staff. Collectively our content-area test development staff members have successfully provided co-facilitation support to PDE for the PSSA Content Item Review meetings, Item/Data Review Committee meetings, and the Bias, Sensitivity,

Fairness Committee meetings since 2004. We look forward to providing our expertise for the GCA review meetings.

Prepare Written Summary Reports of the Content Item Review Meetings

Committees will be asked to accept, accept with revisions, or reject passages and items. DRC may rewrite rejected items during the review process and re-present them to the Committees to maximize item acceptance and to provide an adequate number of items to populate test forms. As stated, DRC will record and document all edits and revisions suggested by committee members. Following the meeting, DRC's content-area test development team members will prepare a final summary report of the results of the meeting. DRC proposes that the summary report will describe the process used during the Content Item Review Committee meetings, number of items that were rejected for not matching the standard, the number of items to be revised and/or were revised during the meetings, and any additional information as requested by PDE. DRC will work with PDE prior to the first Content Item Review Committee meetings with Pennsylvania educators to determine the format for compiling the feedback and preparing the summary reports. In compiling the feedback, DRC will also review any remaining suggested revisions to items with PDE should any remain.

Security Process for the External Committee Reviews

At the beginning of the Bias, Sensitivity, Fairness Review Committee meeting and the Content Item Review meeting, including the Item/Data Review meeting, each committee reviewer who is invited to attend will be asked to sign a Confidentiality Letter specifying the confidentiality agreement and security regulations. The agreement will also outline the ownership regulations. DRC understands that all work developed under this contract will be the sole property of the Commonwealth of Pennsylvania. DRC also understands that no confidential materials related to the project will be released without PDE's explicit approval.

During the review meeting, items, passages, prompts, and scenarios, will not be left unattended. In other words, DRC facilitators will monitor the security of all items, scenarios, passages, and prompts throughout the entire process. All materials sent to the meeting will be sent through a secured mailing process and have tracking documentation. DRC and WestEd facilitators attending the meeting will oversee the delivery of all materials and the return of all materials. These same members will arrange for shredding bins should any materials need to be shredded. In addition, all materials provided to the external committees will be numbered so that secure materials are collected at the end of each day.

VII.B.8. Bias, Sensitivity and Fairness Committees

Support PDE with the Bias, Sensitivity, Fairness Review Committee Process

The delivery of bias-free high stakes, large-scale standards-based assessments is critical to the success of any assessment program. Our item developers follow the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999). DRC and WestEd employ experienced and competent content members who seek to ensure that items and assessments are accurate in terms of content for the Pennsylvania assessments and free of issues of bias, sensitivity, and fairness. In addition to DRC and WestEd's own internal review of passages, items, tasks, and scenarios, our proposal includes our support for PDE's external bias, sensitivity, and fairness review by a panel of experts.

DRC proposes that the Bias, Sensitivity, Fairness Review Committee meetings will take place in Pennsylvania at our DRC program office in Harrisburg. The Committee members will consist of Pennsylvania residents as well as nationally recognized diversity experts. DRC proposes that the Committee each year will consist of five to ten members. The Bias, Sensitivity, Fairness Review Committee meetings will total three to five days for all GCA courses, depending on year. Please refer to *Appendix 12* for detailed information about the bias review meetings.

DRC understands that PDE currently has a file of possible reviewers and that PDE will establish the Bias, Sensitivity, and Fairness Review Committee. DRC will assist PDE in selecting the committee members should PDE request. For example, if PDE requests, we could work with PDE to suggest names of national diversity experts who have had experience reviewing items for bias, sensitivity, and fairness for other large-scale assessment programs. We have provided this support to PDE in the past, and we have worked collaboratively with PDE to seek to ensure that the Bias, Sensitivity, Fairness Review Committee represents the demographics of the Commonwealth of Pennsylvania. Per the RFP requirements, we will assume costs for the meeting space, travel, lodging, food, and relevant expenses for Bias, Sensitivity, and Fairness Committee members. DRC will provide stipends of \$250.00 per day, plus expenses.

In supporting PDE during the Bias, Sensitivity, Fairness Review Committee meetings, DRC's proposal includes overseeing the following meeting activities:

- Coordinating and establishing all meeting logistics, including hotel procedures, meeting rooms, computers, copier capability, etc.
- Assisting PDE with the establishment of the Bias, Sensitivity, Fairness Review Committee membership should PDE request.
- Corresponding with committee members, including information about travel arrangements, meeting announcements, etc., should PDE request. (Note: All written communication will be reviewed and approved by PDE prior to being sent to review committee members.)

- Paying expenses and stipend for committee members.
- Obtaining all supplies, including nametags, travel reimbursement forms, security/confidentiality documents, review tracking documents and approval/revision tally sheets, and other documents, as required by the program.
- Preparing copies of passages and copies of all items, scenarios, and writing prompts. Items, passages, scenarios, and prompts will be printed one per page and banded together by content area unless otherwise directed by PDE.
- Supporting PDE staff, facilitating the review, including compiling the feedback information.

Should PDE request, DRC would be pleased to provide additional support to PDE with the development of Bias, Sensitivity, Fairness Review training materials, including PowerPoint training presentation slides. We will include training in three common areas of bias-related concerns:

- **Opportunity and Access:** The content of the text or test item(s) and task(s) will provide students with a fair opportunity to demonstrate what they know, regardless of their race, ethnicity, gender, religion, disability, socioeconomic status, or region in which they live.
- **Portrayal of Groups Represented:** Issues and/or themes will be approached in a manner that does not demean, offend, or inaccurately portray any religious, ethnic, cultural, gender, social, or disability group.
- **Protecting Privacy and Avoiding Offensive Content:** The content of the text or test item(s), prompts(s), etc., will not intrude on the privacy of the values and beliefs of students or their families, or offend students, parents, or the general public of Pennsylvania.

DRC will support PDE in the facilitation process—a process that will use a consensus method that is designed to encourage all committee members to actively participate throughout the training session and during the review meetings. The reviewers will be assigned to review all content-area passages, items, graphics, scenarios, prompts, etc. Committee members will be encouraged to share their suggestions, ideas, and contributions, as this is a collaborative effort. Each reviewer will record any potential type of bias/sensitivity concern found on the Bias, Sensitivity, Fairness review tracking form with very specific references so that there is no guess-work on the part of PDE or DRC when DRC is compiling the comments.

Proposed Staff for the Bias, Sensitivity, and Fairness Review Committee Meeting

DRC has extensive experience in the facilitation of Bias, Sensitivity, and Fairness Review Committee meetings, both internally as a part of our item review process and externally with bias, sensitivity, and fairness state-specific review panels. We

understand the importance of providing the Commonwealth of Pennsylvania with experienced and appropriate staff, staff members who have had previous experience facilitating fairness and sensitivity review meetings. We propose that Ms. Kimberly Fountain will provide PDE with support for facilitation of these meetings. Ms. Fountain has provided this service to PDE for a number of years. She has facilitated many bias, fairness, and sensitivity reviews throughout the country and provided training for such reviews.

Prepare Written Summary Report of the Bias, Sensitivity, Fairness Review Meeting

The DRC staff members supporting PDE will record all comments. Following the meeting, DRC will compile all feedback and prepare a summary report of the results of the meeting. The summary report will describe the committee review process; the number of items, scenarios, passages, etc., that were rejected for potential bias, sensitivity, and/or fairness concerns; the number of items, scenarios, etc., needing to be revised to remove the potential area(s) of concern; and any additional information as requested by PDE.

Revise Items and Conduct Internal Review Processes

After the external Content Item Review Committee reviews and the Bias, Sensitivity, Fairness Committee reviews have been completed, DRC content specialists will meet with PDE to update the status of the items, prompts, scenarios, etc., as accepted, accepted with revisions, or rejected. All PDE-requested and approved revisions will be made.

To ensure quality of the items prompts, scenarios, etc., and to ensure that all revisions are made during each step in the process, DRC's test development editing team will be responsible for coordinating word-for-word proofreading. At least two editors will perform two independent word-for-word reviews of passages and items to ensure that all requested revisions have been made.

VII.B.9. Statistical and Psychometric Analysis

DRC psychometric staff will be providing all needed analysis of items for the GCA. This analysis encompasses items used both in field testing and items used operationally. DRC psychometrics staff will work with DRC's test development staff to coordinate item analysis and forms construction. Statistical and psychometric analyses go through many phases in a testing program. DRC will analyze all items prior to being placed on forms using the methodologies described below.

Classic Item Analyses (Overall and by Subgroup where Requested)	
<i>p</i> -values, with flags for very easy and very difficult items	Percent choosing each multiple-choice (MC) option, with flags for distractor percent higher than correct-answer percent
Corrected item-total correlations, with flags for possible mis-key or poor item quality (point-biserial)	Option-total correlations for MC items
Percent of students earning each open-ended (OE) item score	Standard error of measurement for the scale
Test reliability	
Differential Item Functioning (for Field Test Items)	
Focal group designation	Reference group designation
Favored group designation	Mantel-Haenszel chi-square for MC items
Mantel-Haenszel delta for MC items	ETS DIF category (A, B, C) for MC items
Standardized Mean Difference (SMD) for OE items	Effect size of SMD for OE items
Effect size of SMD category (AA, BB, CC) for OE items	
Rasch Statistics	
Item Statistics:	
Logit difficulty estimates	Step parameter estimates for OE items
Standard errors for all parameter estimates	INFIT and OUTFIT statistics
Test Indices:	
Test information function	Test characteristic curves
Raw-Logit-scale score tables	Standard errors for all parameter estimates and scale scores
Person separation reliability	

Classical Item Analyses

Field test item analyses will be conducted by form to ensure problems in one form are not masked by other forms. This begins with classical item analysis. DRC uses our proprietary Item and Test Evaluation Modules (iTEMs) system. DRC’s psychometric staff will start with the key verification module of this system; computing the number and proportion of students selecting each response option, the *p*-value for the item, the item-total correlation for the key, and the item-total correlations for each of the response alternatives. These statistics are used to flag any potential incorrect scoring keys. Typically DRC flags items as possibly mis-keyed if the following conditions are observed:

- Percent correct (*p*-value) is low;
- Percent of students selecting any distractor is high;

- Point-biserial correlation for the key is low or negative;
- Point-biserial correlation for a distractor is high.

With iTEMs, the criteria for flagging an item are customizable. As an example, the “low” p -value threshold could be set at any value (e.g., 0.30, 0.35, 0.40). DRC psychometricians will work with PDE to define the criteria that are most suited for the GCA.

Key Verification

The following figure presents a copy of an on-screen display of the iTEMs key verification module. The red flag designates an item flagged that met one or more of the above mentioned criteria.

Item Descriptives(OF)
Core MC Item Detail

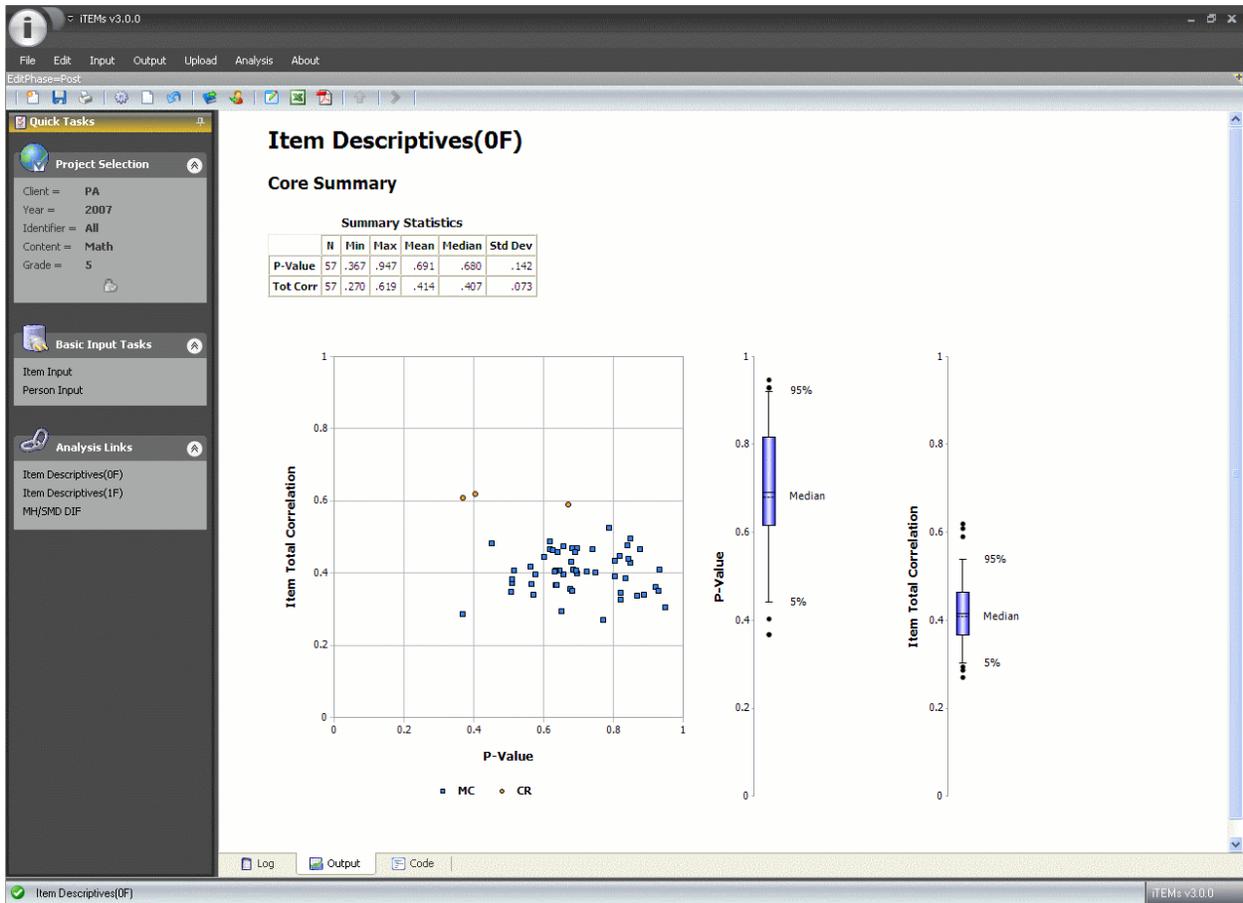
```

/Ability=All
/Ability Correction=True
/Distractor Correlation Method=PtBis
/Item Flags=True
+P-Value<0.3
+Dist Prop>Key Prop
+Dist Prop>0.3
+Item Tot Corr<0
+Item Tot Corr<0.25
+Dist Pt Bis>0
+Dist Pt Bis>Key Pt Bis
    
```

Flag	Id	Seq	Key	#Opts	Wgt	Variables			Frequencies						Proportions						Corre				
						Rept	Cat	Status	Type	II	a	b	c	d	-	*	P-Value	Std Dev	a	b	c	d	-	*	Tot Corr
	41-7080	1	c	4	1	A	O	MC	131709	5912	8799	111787	5148	60	3	.849	.358	.045	.067	.849	.039	.000	.000	-.428	-.231
	41-7077	2	d	4	1	A	O	MC	131709	23564	1182	8641	98277	25	20	.746	.435	.179	.009	.066	.746	.000	.000	-.402	-.357
	41-7046	3	c	4	1	A	O	MC	131709	6697	31798	85735	7405	59	15	.651	.477	.051	.241	.651	.056	.000	.000	-.294	-.147
	41-0268	4	c	4	1	A	O	MC	131709	7147	19936	90106	14451	52	17	.684	.465	.054	.151	.684	.110	.000	.000	-.409	-.189
■	41-7086	6	c	4	1	B	A	MC	131709	4642	25652	48386	52906	102	21	.367	.482	.035	.195	.367	.402	.001	.000	-.286	-.215
	51-2982	7	b	4	1	A	O	MC	131709	1730	124714	3751	1432	63	19	.947	.224	.013	.947	.028	.011	.000	.000	-.305	-.197
	41-7102	8	c	4	1	D	A	MC	131709	13694	5350	81251	31296	78	40	.617	.486	.104	.041	.617	.238	.001	.000	-.487	-.121
	41-0565	9	a	4	1	C	A	MC	131709	105575	1351	2573	22077	88	45	.802	.399	.802	.010	.020	.168	.001	.000	-.434	-.434
	41-7076	10	d	4	1	A	O	MC	131709	5475	12098	16547	97426	117	46	.740	.439	.042	.092	.126	.740	.001	.000	-.467	-.305
	51-2959	11	a	4	1	A	O	MC	131709	101381	8303	3037	18861	99	28	.770	.421	.770	.063	.023	.143	.001	.000	-.270	.270
	51-3119	12	a	4	1	A	O	MC	131709	86391	33193	6376	5615	107	27	.656	.475	.656	.252	.048	.043	.001	.000	-.396	.396
	41-7099	13	b	4	1	B	O	MC	131709	5387	114006	7262	4908	113	33	.866	.341	.041	.866	.055	.037	.001	.000	-.339	-.214
	51-1559	14	d	4	1	E	O	MC	131709	1665	10589	13704	105577	133	41	.802	.399	.013	.080	.104	.802	.001	.000	-.391	-.148
■	41-0023	15	d	4	1	D	A	MC	131709	56829	5481	2161	67104	91	43	.509	.500	.431	.042	.016	.509	.001	.000	-.371	-.240
	41-7002	16	c	4	1	A	O	MC	131709	18689	10208	89212	13433	148	19	.677	.467	.142	.078	.677	.102	.001	.000	-.432	-.198
■	51-2726	17	a	4	1	A	O	MC	131709	59332	3701	9112	59407	116	41	.450	.498	.450	.028	.069	.451	.001	.000	-.484	-.484
	53-5144	18	d	4	1	E	O	MC	131709	23482	4526	11926	91562	171	42	.695	.460	.178	.034	.091	.695	.001	.000	-.469	-.297
	51-2902	19	a	4	1	A	O	MC	131709	116696	7864	1816	5165	134	34	.886	.318	.886	.060	.014	.039	.001	.000	-.341	.341
	41-0021	20	d	4	1	D	O	MC	131709	7799	8573	7562	107552	189	34	.817	.387	.059	.065	.057	.817	.001	.000	-.447	-.245
	41-0506	21	c	4	1	C	O	MC	131709	22043	11201	95083	3201	168	13	.722	.448	.167	.085	.722	.024	.001	.000	-.403	-.282

iTEMs Key Verification Module

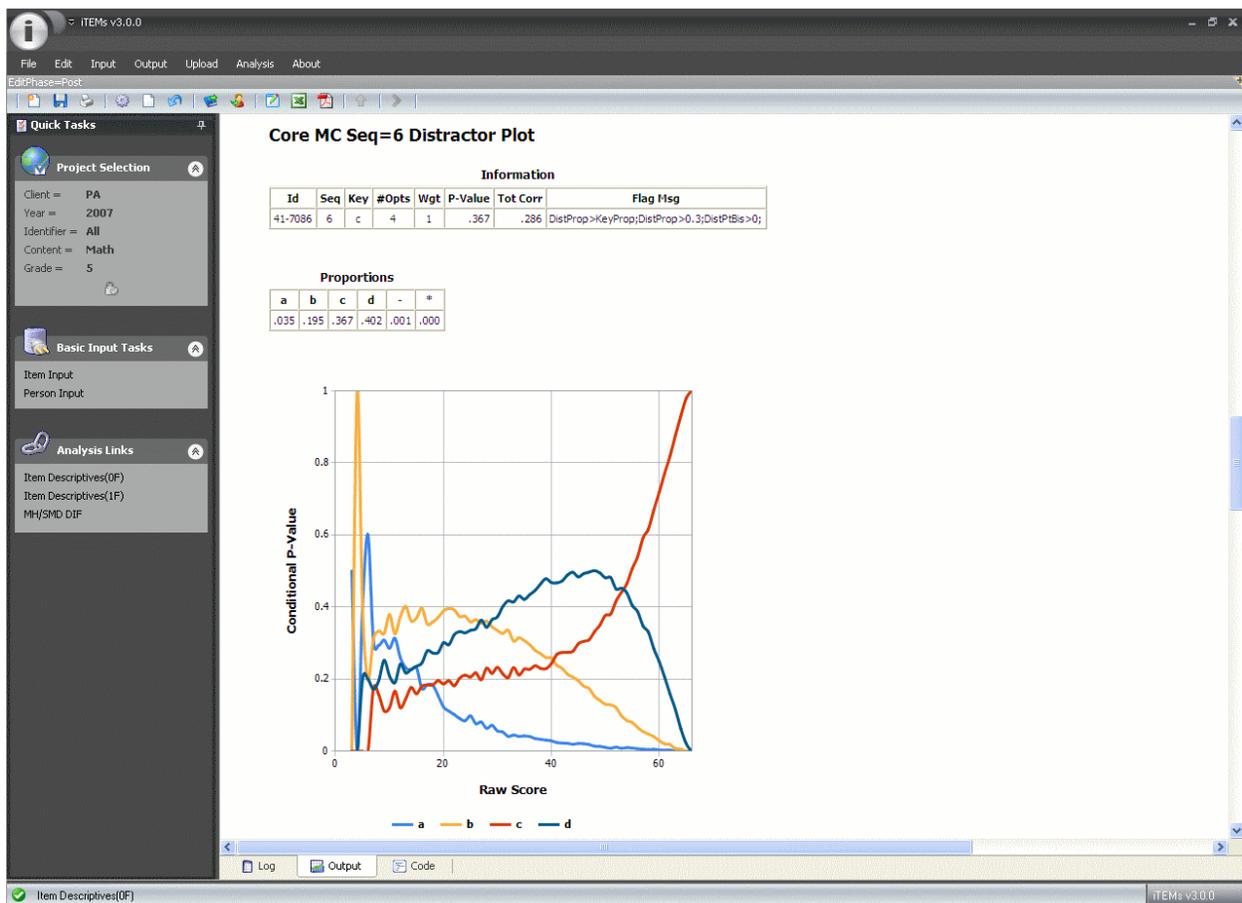
The figure below shows a display of the iTEMs classical analysis module. It presents the p -values and item-total correlations of items allowing for the visual detection of outliers and any other unexpected relationship.



iTEMs Classical Analysis Module

Distractor Analysis

In the distractor analysis, shown in the next figure, iTEMs generates a graph depicting the proportion of students selecting each response option as a function of raw score. The proportion of students selecting the keyed response option should increase as a function of ability (raw score). Conversely, the proportion of students selecting each of the incorrect response options (distractors) should decrease as a function of ability. A graph for an item that does not show this pattern of results may indicate an incorrect key. DRC has found that these item distractor analysis graphs, when used in conjunction with the above-mentioned item statistics, are a powerful tool in detecting possible item mis-keys.



iTEMs Distractor Analysis

The item analysis will be conducted as soon as data based on an appropriate calibration sample are available. **This analysis will be conducted by form.** All items flagged as possibly mis-keyed are immediately referred to DRC Test Development content specialists, Project Management, Information Systems, and Software Quality Assurance staff for further review and verification. Incorrect item keys are identified and evaluated before the final scoring is conducted. Therefore, there are no implications for item calibrations, scaling, equating, and

reporting. Documentation related to any item discrepancies and a copy of the item analysis will be available to PDE for review upon request.

iTEMs can provide p -values, distractor analyses, item-test correlations, percent of students at each open-ended score point, the standard error of measurement, the alpha coefficient, and differential item functioning statistics. In addition, analyses by subgroup can be conducted by gender, ethnicity, ELL status, IEP status, economic disadvantage, and/or other subgroups as requested by PDE.

Differential Item Functioning

DRC will calculate Differential Item Functioning (DIF) statistics to detect possible item bias and these will be reviewed at the *item data review* meetings. DIF analysis is designed to detect items for which students of equal ability from different groups do not have the same probability of answering the item correctly. DIF results will be provided by gender, ethnicity, and other requested subgroupings where sample sizes are sufficient to perform the analyses. Flagged items (i.e., those where the statistical analyses indicate possible DIF) will be reviewed by the Item Data Review Committee.

For multiple-choice items, DRC uses the Mantel-Haenszel (MH) statistic (Mantel & Haenszel, 1959). The MH chi-square test is the most accepted test to determine whether the odds that the *focal* and *reference* groups will respond correctly are equal (Holland & Thayer, 1988; Birch, 1964). The MH analysis involves the computation of a MH chi-square statistic and a MH Delta², or MHD value that represents the average amount that members of the reference group found the studied item more difficult than did comparable members of the focal group. The MH statistic does not depend on the application or the fit of any specific measurement model, does not require specific forms of item response functions, and does not require large sample sizes. This is particularly useful when examining DIF for the GCA based on accommodations, which may have smaller sample sizes.

DRC will report the MH chi-square, delta, and severity classifications. These statistics will enable Pennsylvania educators to make better decisions about the presence or absence of DIF, and will also help to reduce false identification (i.e., labeling items with DIF if no item bias exists).

As an aid to the non-statistical review committees, all items will be placed into DIF severity classifications (*A* +/- to *C* +/-) based on industry standard guidelines (Allen, Carlson & Zelenak, 1999). The *A* category represents negligible DIF, the *B* category indicates moderate potential DIF, while the *C* category indicates that there is large potential DIF. The plus (+) or minus (–) sign that follows the DIF category indicates which group is favored by the item. The minus sign indicates that the reference group outperformed the focal group once the skill level

² The MHD is the ETS delta scale for item difficulty where the natural logarithm of the common odds ratio is multiplied by $-(4/1.7)$.

differences between the groups have been removed. The plus sign indicates that the focal group outperformed the reference group once the skill level differences between the groups have been removed.

A classification system is used in place of a formal significance test of DIF. $MHD < 1.0$ implies it is not significantly different from 0 (based on $\alpha=.05$). The dichotomous DIF classification is defined by the absolute value of $MH \Delta$ squared:

Rule 1: If $|MHD| < 1.0$, the items are classified as A.

Rule 2: If $1.0 \leq |MHD| < 1.5$, the items are classified as B.

Rule 3: If $|MHD| \geq 1.5$, the items are classified as C.

DRC will include the results of the Mantel-Haenszel analyses as in the example shown in the table below. The table shows sample output for multiple-choice items from a DIF analysis report for another state project. Note that a “+” bias code indicates the item favored the focal group, while a “-” bias code indicates the item favored the reference group.

Sample Multiple-Choice (MH) DIF Analysis Report

Form	Unique Item ID	Reference Group	Focal Group	MH Chi Square	MH Delta	DIF Classification
A	5546	male	female	15.41	1.0	A+
A	2654	male	female	44.14	1.2	B+
A	2187	male	female	18.56	0.6	A+
A	2277	male	female	48.46	1.5	B+
A	2364	male	female	15.92	0.8	A+

The analysis of open-ended (OE) items will be based on the *standardized mean difference* (SMD) procedure developed by Zwick and Thayer (1996), which takes into account the natural ordering of the response levels of the item. In contrast to the MH procedure, this summary statistic compares the means of the reference and focal groups, adjusting for differences in the distribution of each group’s members across a given number of ability (i.e., total score) stratifications. The SMD statistic represents the difference between the unweighted item mean of the focal group and the weighted item mean of the reference group. The weights applied to the reference group are determined such that the weighted number of reference group students is the same as in the focal group (within the same ability group). The SMD is then divided by the total group item standard deviation, resulting in a measure of the *effect size* (ES) for the SMD.

Based on this *ES* value, and its statistical significance, all items will be placed into DIF severity classifications (AA+/- to CC+/-) based on NAEP guidelines (Allen, Carlson & Zelenak, 1999). The polytomous DIF classification is defined as:

Rule 1: if $|ES| \leq 0.17$ classify the item as AA³.

Rule 2: If $0.17 < |ES| \leq 0.25$, classify as BB.

Rule 3: If $|ES| > 0.25$, classify as CC.

DRC will include the results of the Mantel-Haenszel and SMD analyses in the Technical Report. Tables for OE items will be displayed in the same format as shown for MC items above.

Using iTEMs for DIF Analysis

DRC's proprietary system, iTEMs, may be used to present DIF analyses. With iTEMs, the ability grouping function is fully customizable. The groupings can be made by using either a given number of groups (fixed) or an overall group minimum (dynamic) as an ability grouping rule. Items can be flagged based on the following criteria:

- DIF code
- Overall ability group minimum
- Reference ability group minimum
- Focal ability group minimum

³ $ES < 0.17$ is not significantly different from 0 (based on $\alpha=.05$)

The figure below presents a screenshot of the iTEMs DIF module. The red flag designates an item flagged that met one or more of the criteria for DIF.

ITEMs v3.0.0

File Edit Input Output Upload Analysis About

EdiRPhase=Post

Output Tests

Exterior MC Item Detail

/Ability=Core
 /Number Of Groups=4
 /Valid Grouping=True
 /Variable=Gender
 /Reference Group=Male
 /Focal Group=Female
 /Item Flags=True
 +Codes=B+;B-;C+;C-;

Flag	Form	Id	Seq	Key	#Opts	Wgt	Variables			MH DIF										Code	Flag Msg	
							Rept Cat	Status	Type	II	Ref II	Foc II	Min	Ref Min	Foc Min	Ref P-Val	Foc P-Val	Chi-Sq	Alpha			Delta
	1	61-1412	5	b	4	1	A	F	MC	7742	4022	3720	1850	922	875	.673	.662	.547	1.040	-.092	A	
	1	51-1240	58	c	4	1	E	M	MC	7742	4022	3720	1850	922	875	.628	.622	.235	1.026	-.061	A	
	1	51-3729	59	b	4	1	D	M	MC	7742	4022	3720	1850	922	875	.664	.697	8.323	.858	.361	A+	
	1	51-2913	60	c	4	1	B	M	MC	7742	4022	3720	1850	922	875	.552	.569	3.134	.914	.212	A	
	1	51-2838	61	a	4	1	A	M	MC	7742	4022	3720	1850	922	875	.692	.638	37.283	1.402	-.795	A-	
	1	51-2040	62	d	4	1	C	M	MC	7742	4022	3720	1850	922	875	.596	.580	2.705	1.088	-.198	A	
	1	61-2138	63	c	4	1	C	F	MC	7742	4022	3720	1850	922	875	.666	.642	5.757	1.132	-.292	A-	
	1	61-1402	64	a	4	1	D	F	MC	7742	4022	3720	1850	922	875	.736	.786	29.046	.704	.824	A+	
	1	61-1420	65	a	4	1	D	F	MC	7742	4022	3720	1850	922	875	.528	.600	57.524	.660	.977	A+	
	1	61-1414	66	d	4	1	A	F	MC	7742	4022	3720	1850	922	875	.421	.446	6.768	.875	.313	A+	
	1	61-2128	67	a	4	1	B	F	MC	7742	4022	3720	1850	922	875	.782	.794	1.204	.935	.157	A	
	1	61-1401	68	c	4	1	E	F	MC	7742	4022	3720	1850	922	875	.837	.878	24.066	.685	.889	A+	
	2	61-1929	5	c	4	1	A	F	MC	7150	3704	3446	1758	840	807	.661	.668	2.090	.925	.183	A	
	2	51-1948	58	a	4	1	C	M	MC	7150	3704	3446	1758	840	807	.772	.766	.003	.995	.011	A	
	2	41-7279	59	b	4	1	E	M	MC	7150	3704	3446	1758	840	807	.690	.621	31.389	1.393	-.780	A-	
	2	51-3679	60	c	4	1	D	M	MC	7150	3704	3446	1758	840	807	.404	.441	14.678	.828	.444	A+	
	2	51-3225	61	b	4	1	A	M	MC	7150	3704	3446	1758	840	807	.725	.594	143.626	2.001	-1.630	C-	Code=C-
	2	41-7297	62	c	4	1	A	M	MC	7150	3704	3446	1758	840	807	.756	.800	28.967	.721	.770	A+	
	2	61-1416	63	a	4	1	E	F	MC	7150	3704	3446	1758	840	807	.726	.753	14.161	.801	.520	A+	
	2	61-1419	64	a	4	1	D	F	MC	7150	3704	3446	1758	840	807	.615	.671	51.832	.659	.981	A+	
	2	61-2297	65	b	4	1	A	F	MC	7150	3704	3446	1758	840	807	.886	.893	1.626	.896	.258	A	
	2	61-1318	66	a	4	1	B	F	MC	7150	3704	3446	1758	840	807	.620	.535	45.045	1.468	-.902	A-	
	2	61-1711	67	c	4	1	D	F	MC	7150	3704	3446	1758	840	807	.660	.695	22.099	.766	.625	A+	
	2	61-2032	68	a	4	1	C	F	MC	7150	3704	3446	1758	840	807	.707	.712	2.534	.901	.244	A	
	3	61-1929	5	c	4	1	A	F	MC	7133	3619	3514	1695	803	851	.667	.672	.193	.976	.058	A	
	3	51-1095	58	b	4	1	D	M	MC	7133	3619	3514	1695	803	851	.327	.347	4.972	.888	.280	A+	

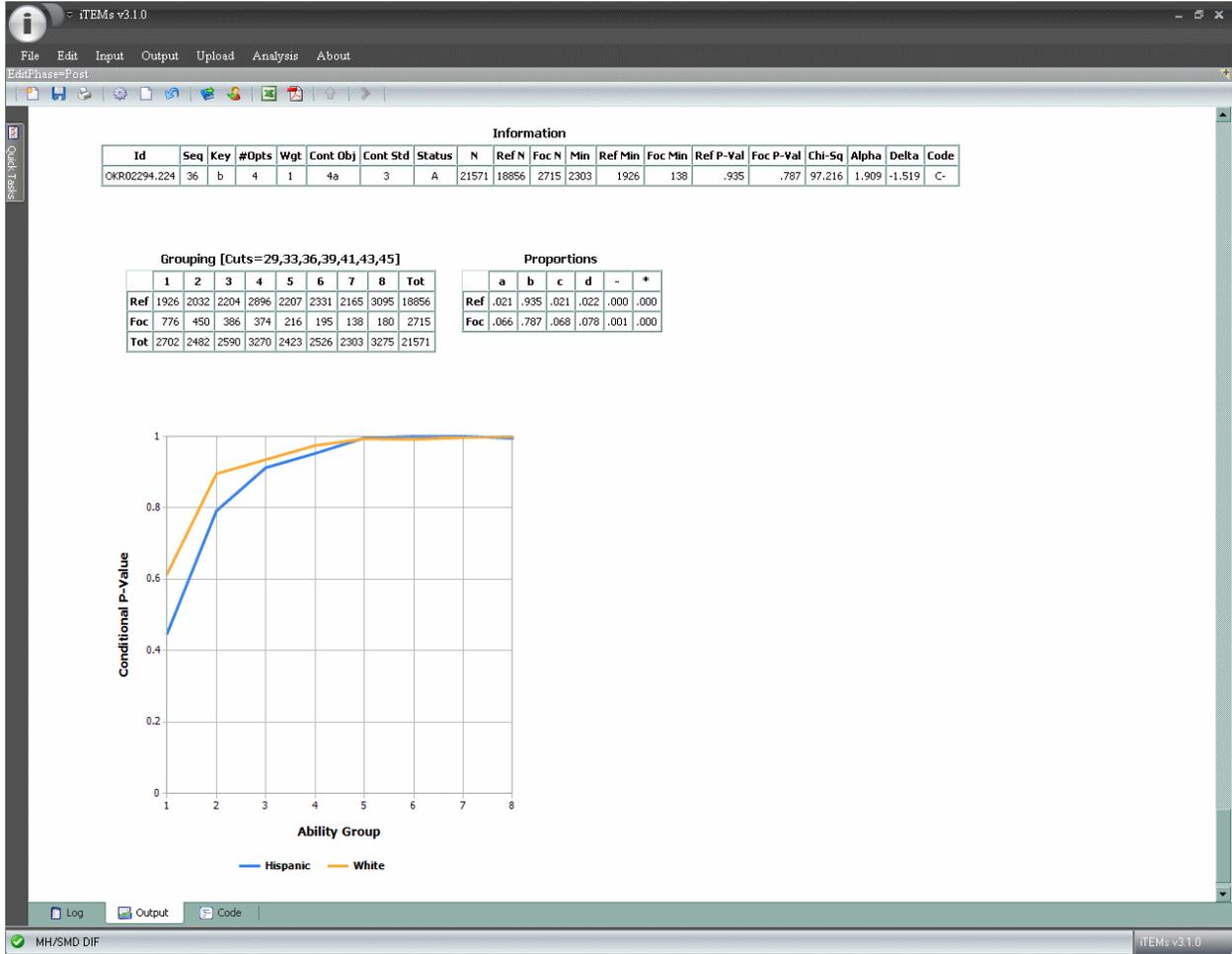
Log Output Code

MH/SMD DIF

ITEMs v3.0.0

iTEMs DIF Module

Below is an example of a multiple-choice item plot in *iTEMs* of the expected performance of males and females.



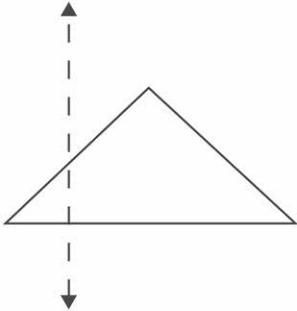
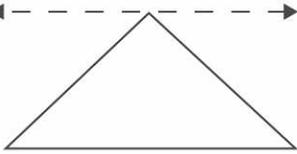
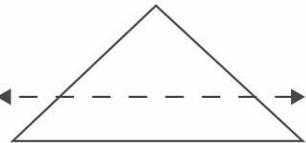
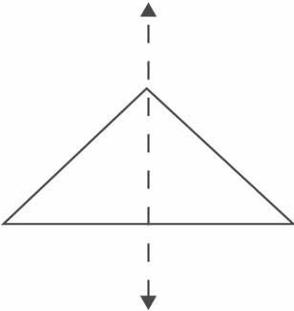
DIF Item Plot

Item Data Review of Field Test Items

In addition to item reviews for newly developed items—which are focused on content and bias and sensitivity—DRC will facilitate a *data review* of field tested items. This review will be conducted after the items have been field tested and prior to selection of items for the operational administration. It will focus on content validity, curricular alignment and the statistical functioning of newly-developed items. DRC proposes the same types of support for the item with data review meetings, as discussed under *Subheading VII.B.7*. Please refer to *Appendix 12* for more information about the item and data review meetings.

The Pennsylvania educators must meet minimum qualification standards as determined by PDE. DRC will cover educator travel, lodging, food, and relevant expenses.

The review will be co-facilitated by DRC content assessment specialists and DRC psychometricians. The goal of this review is to ensure that only high-quality items are made available for the construction of the base forms for upcoming testing cycles. Item level data will be brought to the Item Data Review on item data cards, which will include statistics from the field test analysis. An example can be seen in the following figure. Item cards will contain answer keys or rubrics and associated data with the complete item information. A sample item card is presented in the following figure.

Standard: Identify/draw one line of symmetry in a two-dimensional figure.		GCA New Item Review
1. Monica has a part of a grilled cheese sandwich. Her part of the sandwich is shaped like a triangle. Which picture shows a line of symmetry on Monica's part of the sandwich?		Item ID
<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>A. </p> </div> <div style="text-align: center;"> <p>B. </p> </div> </div> <div style="display: flex; justify-content: space-around; margin-top: 20px;"> <div style="text-align: center;"> <p>C. </p> </div> <div style="text-align: center;"> <p>D. </p> </div> </div>		520101
		Content Area
		Math
		Grade Level
		3
		Rpt Category
		C
		Asmt Anchor
		2
		Sub-Anchor
		1
		Eligible Cont.
		1
		Primary Code
		C.2.1.1
		Item Type
		MC
		Points
		1
		Depth Knldg
		1
		Est Difficulty
		Easy
		Answer Key
		D
		Calc Use
		NC
		Focus
		Symmetry

Administration										Fit Statistics			
Name	Use Function	Seq	Period	Year	Day	Session	Calc	Model/Ext	Grade	Outfit t	Infit t	Deg Free	Chi-sq/df
M03F	FT	40	Spring	2007		1	No	RASCH	3	-1.1	-1.4		

Traditional Statistics				IRT Statistics				
# in group	P-Value	Item Mean	Item/Tot Corr	Label	Final	Final S.E.	Preliminary	Preliminary S.E.
1962	85.22		0.41	Location	-1.85	0.07		

Distractor/Step Specific				DIF Analysis		
Label	Percent	Correlation	Step Logit	Category	Bias Code	Num Value
A	2.29	-0.16	-0.90	Male/Female	A-	-0.15
B	8.51	-0.26	-0.69	White/Hispanic	A-	-0.89
C	3.98	-0.25	-1.03	Lunch/Aid	A-	-0.79
D*	85.22	0.41	0.63			

A: verticle line, not line of symmetry
 B: horizontal line, touches top of triangle
 C: horizontal line, across middle of triangle
 D: correct

Sample Item Card

DRC Psychometric Services staff will provide committee members with the statistical training necessary to make appropriate, well-informed, item-related decisions. This training will help the committee members understand how both statistics and content are needed to present a clearer representation of an item's performance than may be gained by either statistical or content review alone. To provide the appropriate level of psychometric support, senior staff from DRC's Psychometric Services Department will oversee the training. During training, and subsequent data review, the following topics will be discussed.

- p -values (percent correct)
- Point-biserial correlations
- Role of item alternatives
- Analysis of item response patterns
- Differential Item Functioning (DIF): introduction to bias-free measurement
- Potential effects of location of the item on the form
- Number of students who omitted an item
- Data on the curricular alignment of the items
- Other materials as may be necessary to review the functioning of the items

The psychometric data play a critical role in determining whether an item eventually is accepted into the operational item bank; however, they should not be the sole determining factor. Psychometric data, for any given item, must be viewed as cautionary flags to draw attention to potentially problematic items. Whether such items are ultimately approved for use on an operational form will then depend on content and curricular considerations. The fact that these sessions will be co-facilitated by both content and psychometric staff underscores this important balance. DRC content assessment staff will address this and other topics during training.

Committee members, with the assistance of DRC Psychometrics staff, will consider the statistical *quality* of the items. The specific triggers for flagging items will be guided by PDE. The following are examples of various item flagging criteria:

- Item means less than half the points possible for open-ended (OE) items.
- Item-to-total correlations less than .20.
- Option-to-total correlations for incorrect multiple choice (MC) answers that are greater than 0.00.
- Attractiveness of all answer choices for MC items.
- Expected patterns of percent earning each score based on the overall difficulty of the item for OE items.

- Differential Item Functioning codes of B or C.
- Poor Rasch fit statistics.

Based on the consensus of the committee members, items will be accepted, accepted with minor revisions, or rejected for use on operational forms. Items that are rejected by the committee, after collaboration with PDE, can either be revised and re-field tested where appropriate, or removed from the item bank.

DRC will maintain a complete record of all relevant committee actions, recommendations, comments, and rationale. DRC will consult with PDE on all decisions regarding the status of the items. DRC acknowledges that PDE will have final approval of all changes made.

After they have been approved by the item review, DIF, and data review committees, items are ready for forms construction.

Rasch Analysis

Field test items will be calibrated using Rasch methodology. For a detailed description of Rasch calibrating, please see *Subheading VII.J*.

VII.C. ASSESSMENT CONSTRUCTION

VII.C.1. Construction of Assessment Forms

As discussed above, DRC agrees to support, provide technical assistance, and work cooperatively with PDE content and assessment specialists to select passages, scenarios, and prompts with associated content-specific items for the GCA. DRC will construct forms that comply with the test blueprints and test specifications. As discussed above, and in more detail in sections that follow, DRC's test design includes a core or common set of items, along with embedded field test items. Individual student scores will be based on their responses to the core/common items only. DRC has extensive experience with this test design, and will work with PDE to implement.

DRC has successfully used and endorses an integrated team approach to test construction, including content area specialists, psychometricians, and scoring specialists working as a unit in collaboration with client development teams. DRC experience has shown that, in the case of collaborative efforts between integrated teams of developers, the whole certainly becomes more than the sum of the parts.

DRC submits the following test construction plan for PDE's consideration. All aspects specified in the RFP are included. It is understood that, simply because of the nature of the task, modifications to this plan may well be necessary and DRC is prepared to make adjustments and adaptations as required. DRC will follow the test construction plan regardless of the mode of delivery—paper-and-pencil or online.

Forms Building

As a first step in building the GCA assessments, DRC will create, store, and maintain all potentially usable items in IDEAS. The form, format, extent, and organization of items will be determined in consultation with PDE. DRC acknowledges that PDE will expect an item to include statistical and meta-data along with scoring guidelines. Related graphics, tables, and charts will also be provided. The intent, with direction from PDE, is to organize test construction material and resources, including items with associated graphics and data, in as logical and user-friendly form as possible.

Following preparation of all necessary material and resources, forms construction will begin. Construction of the test forms themselves will be a collaborative effort between PDE and DRC's integrated development team of assessment specialists, psychometric services specialists, and scoring specialists.

DRC understands that before test forms are created, passages, items/performance tasks, scenarios, and artwork must be carefully selected. Below, we have described the process used for item selection; however, we are open and ready to refine our process to best meet the needs of PDE.

Process for Selecting Items for Forms

1. Using the pool of items approved by Pennsylvania educator committees, DRC test development specialists will first select items to match the approved test blueprints.
2. DRC test development specialists will check to see that each item clearly aligns with anchor and /or content standards where applicable, and that each item meets psychometric guidelines for excellence.
3. DRC test development specialists will verify that each item meets technical quality for well-crafted items, including:
 - One clearly correct answer
 - Clear and concise wording
 - Grammatical correctness
 - Appropriate range of difficulty
 - Free of any offensive, inappropriate, or biased content
 - Meets the principles of universal design and maximum accessibility.

Once DRC test development specialists receive PDE approval on the items to be placed in test forms, test form construction will begin. In constructing the forms, DRC content area test development specialists will follow the guidelines provided below.

Some Guidelines for Placing Items into Forms

- Forms will include adequate standards coverage, as required by test blueprints.
- No item in a form will “clue” another item on that same form.
- “Clang” will be avoided (i.e., distractors should be unique from one another).
- Forms will be ethnically diverse as needed, in terms of artwork and graphics.
- Forms will include a wide range of topics and a variety of questions.
- Correct answer distributions will be psychometrically sound.
- Forms will not contain any items that have been released to the public.
- PDE will review and give final approval of all test forms.

DRC psychometricians will examine the statistical quality of preliminary test forms, paying specific attention to:

- Targets
 - Test Information Function
 - Test Characteristic Curves
- p -value
- Fit Statistics
- Potential Item Bias
- Key Distribution

Any items found to be suspect from a statistical standpoint will be reported to the test development content specialists for review and possible replacement. This process is repeated until a form satisfies both internal groups and PDE. DRC believes that test form construction is a collaborative effort. Our test development specialists build a test form based on the content distribution described by the test specifications and blueprints determined by PDE. We pay particular attention to ensuring that potential problems related to developmental appropriateness, item cueing, or redundant content are eliminated. DRC utilizes the process described below to construct forms.

Forms Construction Process

1. Test development content specialists review the approved *Pennsylvania's Assessment Anchor Content Standards* and/or the *Pennsylvania Academic Standards*, curriculum frameworks (big ideas, concepts, and competencies), test specifications, and test blueprints.
2. Psychometricians provide test development content specialists with an overview of the psychometric guidelines for forms construction, ensuring that our process is approved by PDE.
3. Test development content specialists and measurement experts are trained in forms construction, with a focus upon requisite content validity and psychometric properties.
4. Test development content specialists receive all items and accompanying data and are trained to use the form-building software to build forms.
5. Test development content specialists review all items in the operational pool and make an initial selection of anchor items according to test blueprint guidelines and psychometric guidelines, such as item fit statistics and the stability over time of the proposed linking items.
6. Test development content specialists review linking item selection, following the guidelines for meeting psychometric and content technical quality.
7. Test development content specialists will work with PDE to make replacements, if needed.
8. Test development content specialists make an initial selection of items according to test blueprint and psychometric guidelines and review non-linking item selections for psychometric and content technical quality.
9. Test development content specialists create item-mapping charts for each test.
10. Items selected for forms construction, with item mapping charts, will be reviewed by DRC senior-level test development content specialists and measurement experts.
11. Final recommendations for items selected for forms construction will then be prepared for review by PDE along with non-anchor selections within forms.
12. Test development content specialists will make suggested replacements.
13. Final sign-off will occur between DRC and PDE.

DRC understands that PDE will have final approval of the selection of items and test forms, and we agree to work cooperatively to ensure a smooth flow of information between PDE's assessment specialists and DRC's test development team. We commit to quick turn around of edits to ensure timelines are met and that the program remains on schedule.

Test Build—Using FirstForm for the GCAs

DRC proposes using FirstForm, DRC's proprietary form building application. DRC can use FirstForm for either manual or automated forms construction. In manual mode, it provides the ultimate flexibility and ease of use for both content specialists and psychometricians. In automated mode, preliminary forms can be constructed based on a list of supplied constraints before they are reviewed by content specialists. The former mode works best when the item pool is relatively small, whereas the latter mode can significantly reduce the time required for building forms when the item pool is relatively large. In either mode, electronic documentation is stored and can be downloaded and printed for formal sign-off.

FirstForm features flexible constraint handling, modeled after industry best practices (e.g., Fletcher, 2000). Classes of constraints are listed below.

FirstForm Classes of Constraints

- Statistical objective functions, such as maximizing the test information function near specific ability cutoff points, minimizing deviation of the test characteristic curve from the target test characteristic curve, and simplifying the construction of multiple parallel forms.
- Content distribution requirements within each of several content strands or within combinations of content strands.
- Total number of multiple-choice items and points.
- Item dependencies, such as positive dependencies (friend-relation) between a multiple-choice item and a writing prompt and negative dependencies (enemy-relation) between items that are too similar to reside on the same form or that might act as semantic primers or cues in relation to one another.
- Number of items using specific materials such as reference sheets.
- Minimum and maximum point-biserial correlations, p-values, IRT item difficulty parameters, and item discrimination.
- Forced or required items such as anchor items or linking items.
- Inclusion or exclusion of items with specific administration history, such as eliminating items from a specific field test or operational administration, limiting the pool only to those items from a set of field test administrations, or using an upper bound on the number of times an item may be administered.

Some uses of FirstForm may include:

- **Manual Assembly:** A test form can be constructed manually by selecting items one by one or in groups or testlets directly from a pool of candidate items. Items can be sorted and filtered just as in a spreadsheet application. The user can also apply the constraint filters that are available with automated test assembly.
- **Linear Program Assembly:** Test forms can be constructed from large item pools under a sizable set of constraints using linear programs. The software tool translates user requirements into a set of equations to be solved by standard linear programming algorithms.

Below is an example of how constraints may be entered into FirstForm.

Target Definition

Use existing item set as target >> [Dropdown Menu]

Enter a Rasch Logit to define a target curve >> [Input Field]

Enter 3 IRT parameters to define a target curve >> [Input Field] [Input Field] [Input Field]

b = location parameter
a = slope parameter
c = lower asymptote

Enter a P-Value to be used as target difficulty >> [Input Field]

Cancel Finish

The next two figures show summary-level data once a form is constructed. Notice in the first example that for the summary data, on the Form Browser, choices include Form and Strand detail, Curve, and Item Swap. In the second example, a more in-depth item summary is provided.

Form Browser

Form

Strand Detail

Curve

Form Detail

Item Swap

Content Summary

Content	Count	Percent
DC.0	0	0.0%
DC.2	0	0.0%
DC.3	0	0.0%
DC.4	0	0.0%
MT.2.A	0	0.0%
MT.2.B	0	0.0%
MT.2.C	0	0.0%
MT.2.D	0	0.0%
NS.0	0	0.0%
NS.1.A	1	2.5%
NS.1.B	0	0.0%
NS.1.C	1	2.5%
NS.1.D	0	0.0%
NS.1.E	0	0.0%
NS.2.A	0	0.0%
NS.2.B	3	7.5%
NS.2.C	0	0.0%
NS.2.D	0	0.0%
NS.3.A	5	12.5%
NS.3.B	9	22.5%
NS.3.C	1	2.5%
NS.3.D	0	0.0%
NS.3.E	0	0.0%
SS.0	1	2.5%
SS.1.A	3	7.5%

Location Summary

Stat	Value
Mean	0.20
SD	0.23
Min	-0.16
Max	1.08

P-Value Summary (MC)

Stat	Value
Mean	76.88
SD	4.73
Min	62.00
Max	83.00

Point Biserial Summary (MC)

Stat	Value
Mean	0.39
SD	0.08
Min	0.14
Max	0.53

Slope Summary

Stat	Value
Mean	0.00
SD	0.00
Min	10000000.00
Max	10000000.00

Mean Response Summary (CR)

Stat	Value
Mean	0.00
SD	0.00
Min	0.00
Max	0.00

Item-T total Correlation (CR)

Stat	Value
Mean	0.00
SD	0.00
Min	0.00
Max	0.00

Asymptote Summary

Stat	Value
Mean	0.00
SD	0.00
Min	10000000.00
Max	10000000.00

Answer Key Summary

Key	N	N%
1	8	20.0%
2	15	37.5%
3	10	25.0%
4	7	17.5%
TOTAL	40	100.0%

Model Parameters

Test Characteristic Curve

Test Information Function

Number of Items: Number of Strands:

Item Parameters - Location

Mean: SD:

Item Parameters - Slope

Mean: SD:

Item Parameters - Asymptote

Mean: SD:

Item Parameters - MC P-Value

Mean: SD:

Item Parameters - CR Mean Response

Mean: SD:

Item Parameters - Point Biserial

Mean: SD:

Export this Form to Excel Edit this Form Exit and Save

Test Information Function

The picture above (the chart in the lower left corner) shows a sample Test Information Function (TIF). The TIF shows how much information is provided at particular scaled scores. The TIF is the reciprocal of the squared standard error of measurement, so the greater the information, the lower the measurement error at a specific location on the score scale. TIFs generally follow an inverted-U-shaped curve as illustrated.

For tests like the GCAs, which have a single 'passing' score, there are benefits in selecting test items that will provide maximum information at the passing score. This will be a goal during DRC's test construction process. Of course, item quality and content representation should never be traded for this degree of psychometric purity. Thus, the goal will be to maximize information at the cut given the constraints of the item pool and the content specifications. This will provide the greatest assurance that the GCA Proficient/Not Proficient classifications will be as reliable as possible.

The following procedure, based on Lord (1977), can be useful when constructing tests using item-information and/or error curves:

1. Decide on the target test information curve.
2. Select items from the item bank with item-information curves that will fill up the hard-to-fill areas under the target-information curve.
3. After each item is added to the test, re-compute the test-information curve for the selected test items.
4. Continue selecting test items until the selected test-information curve matches the target information curve to a satisfactory degree.

In summary, DRC will select items to maximize test information at an important location on the ability scale spanned by the examinees being tested (i.e., the module cut score). This "optimum" selection of test items will contribute substantially to the precision with which ability scores are estimated at each GCA cut score.

Test Characteristic Curves

The picture above (the chart in the upper left corner) shows a sample Test Characteristic Curve (TCC). The TCC defines the relationship between raw scores and scale scores. For most tests, the TCC will follow the familiar S-shaped curve similar to the logistic curve. The slope of the curve is closely related to the test information function with a steeper slope corresponding to greater information and less measurement error at that location of the scale.

If two forms have identical TCC then the forms are "strictly parallel" and require no additional equating. This situation is also a goal of the DRC test construction process for the GCA. The measurement model will allow for some adjustment in

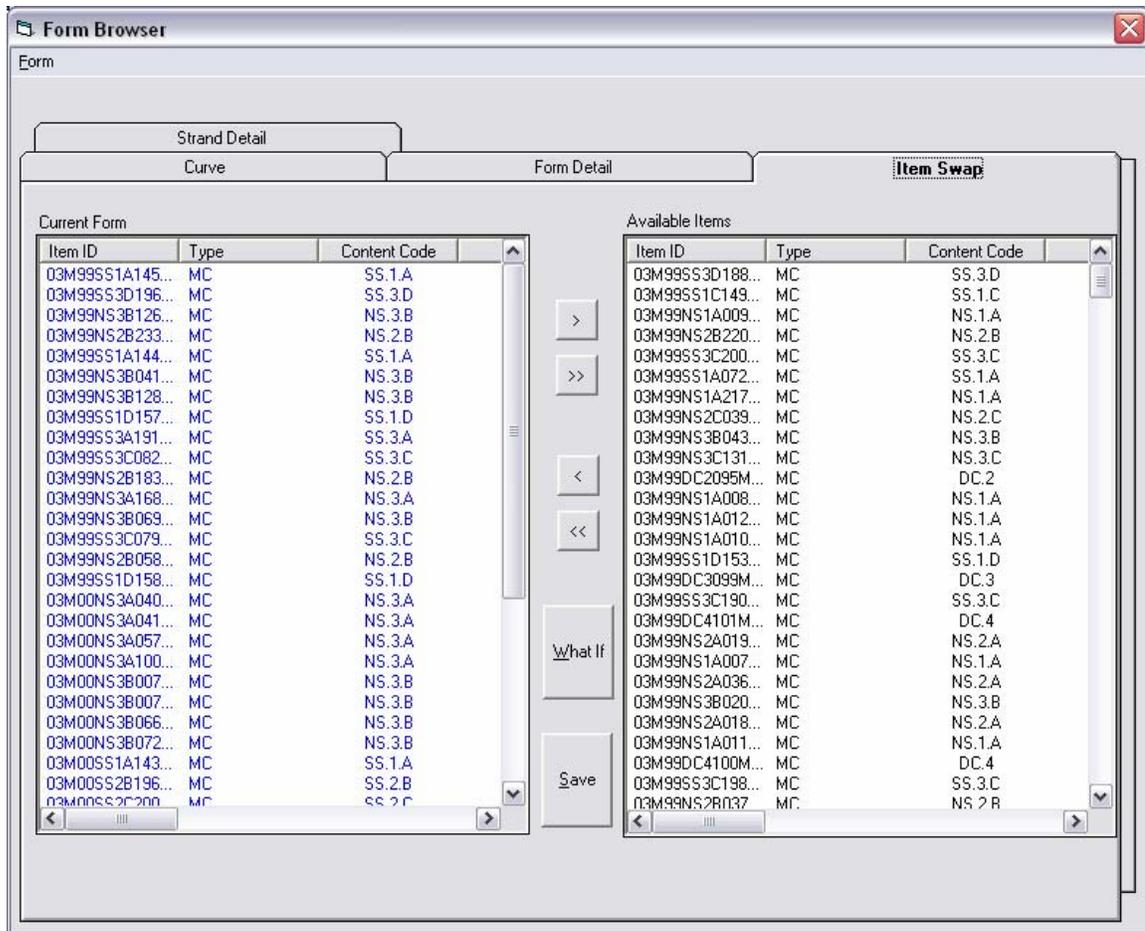
the equating, even though this may result in possible differences in the cut scores of one or two raw score points.

The goal will be to match the TCC as closely as possible to the target given the constraints of the item pool, the content specifications, and the primary goal of maximizing information at each GCA cut score (as described above). This will provide the greatest assurance that the assessment will be consistent across administrations and any differences in student performance represent true changes.

Item Swap

One of the most important features of FirstForm is the “what if” option (see the following figure). This option allows the content specialist to try out different combinations of items on a form. The user produces “what if” plots that display the statistical properties of their trial forms—in particular, the test characteristic curves and information functions, illustrated in the two stacked graphs on the far left side of the previous figure. It is this functionality that can be of the greatest use in the construction of the GCA tests, as it allows test developers to target an area of interest on the ability scale and then construct the test to be maximally informative at that location. This targeting helps to ensure that students are not misclassified as non-proficient due to measurement error. Constructing the GCA tests to *maximize information* at the Proficient cut is a critical proactive step in ensuring that the tests provide the *greatest precision* in the region of the score scale that determines whether students meet the designated standard.





If PDE's review indicates a potential problem with one of more items, the residual pool of operational-ready items is available within the FirstForm system so that replacement items can be chosen. In effect, the new form can be documented and approved on the spot, saving a tremendous amount of time and effort over non-electronic processes.

When a final form is achieved, a simple click of a button produces formal, electronic documentation of the form, including a sign-off sheet for DRC test development specialists, the Lead Psychometrician, and PDE. This feature is particularly attractive as part of the forms construction review process

DRC Test Development Forms Construction Quality Check

After forms construction, Test Development editors and specialists implement quality-control procedures to ensure accuracy of all GCA test forms. The Test Development Team will review test forms in IDEAS. Using original copies previously approved by PDE, they will make modifications, if necessary, to verify continuity and accuracy. Once stable versions of materials have been achieved (i.e., content, graphics or illustrations are complete and accurate), final proofs will be submitted to the editing/proofreading team.

DRC’s Test Development Editorial Team will be responsible for coordinating word-for-word proofreading of all test forms; at least two editors will perform three independent word-for-word reviews. These specialists will proof and will query potential issues in order to offset the possibility of misleading or technically problematic wording of items within test forms. The Editing team will conduct a final “three-way” proofreading (test booklet, answer document, and test directions) of forms and confirm any potential inaccuracy with DRC Test Development Specialists.

In addition, DRC Psychometric Services and Test Development Specialists will also review all test forms to confirm accuracy and continuity; they will continue to query any test item that is suspect from a content standpoint. These content staff will enlist the aid of external reviewers to take each of the tests to verify the correct answers and content standard alignment for the multiple-choice items. They will record their answers to the items on each test to confirm the scoring keys. Psychometric Services and Test Development staff will work with PDE to revise scoring keys whenever necessary and will have these revisions verified and approved by the Software Quality Assurance Analyst and PDE. PDE will be provided final, clean copy for its review and approval to print.

VII.C.2. Initial Field Test Plan

DRC is pleased to present the following detailed plans for initial field testing. DRC proposes a sample field test for all subjects (1,000 students per form) except for English Composition, which DRC proposes as two separate field-test events.

Algebra I, Algebra II, and Geometry

Our initial field test for Algebra I, Algebra II, and Geometry will take place in Spring of 2010. As you can see our plan includes a robust number of items to allow for attrition at data review. In addition, this plan would provide items to build the first 5 operational forms (4 core and 1 breach) from the initial field test for the administrations in Fall (Dec.) 2010, Spring 2011, Summer 2011, and Fall 2011.

Algebra I, Algebra II, and Geometry

Points Plan % MC & OE	Unique FT MC Items per FT Form	Common FT MC Items per Form	Unique FT OE Items per FT Form	Total FT Items per FT Form MC/OE	Total Points per FT Form	Total FT Forms
50/50	30	15	9 4pt 6 2pt	45 MC 15 OE	93	6
65/35	39	15	6 4pt 6 3pt	54 MC 12 OE	96	6
80/20	48	15	6 4pt	63 MC 6 OE	87	6

Biology and Chemistry

Our initial field test for Biology will take place in Spring of 2011. As you can see our plan includes a robust number of items to allow for attrition at data review. In addition, this plan would provide items to build the first 5 operational forms (4 core and 1 breach) from the initial field test for the administrations in Fall (Dec.) 2011, Spring 2012, Summer 2012, and Fall (Dec.) 2012. The same plan will be used for Chemistry with an initial field test in Spring 2011 supplying items for the administrations in Fall (Dec.) 2011, Spring 2012, Summer 2012, and Fall (Dec.) 2012.

Biology and Chemistry

Points Plan % MC & OE	# of Scenarios per Form	Unique FT MC Items per FT Form	Unique Common FT MC Items per Form	Unique FT OE Items per FT Form	Total FT Items per FT Form MC/OE	Total Points per FT Form	Total FT Forms
50/50	4	30	15	9 4pt 6 2pt	45 MC 15 OE	78	8
65/35	4	39	15	6 4pt 6 3pt	54 MC 12 OE	81	8
80/20	4	48	15	6 4pt	63 MC 6 OE	72	8

Literature

Our initial field test for Literature will take place in Spring of 2011. As you can see our plan includes a robust number of items to allow for attrition at data review. In addition, this plan would provide items to build the first 5 operational forms (4 core and 1 breach) from the initial field test for the administrations in Fall (Dec.) 2011, Spring 2012, Summer 2012, and Fall (Dec.) 2012.

Literature

Points Plan % MC & OE	# of Unique Passages per Form	# of Passages per Form	Unique FT MC Items per FT Form	Unique Common FT MC Items per Form	Unique FT OE Items per FT Form	Total FT Items per FT Form MC/OE	Total Points per FT Form	Total FT Forms
50/50	9	12	30	15	9 4pt 6 2pt	30 MC 15 OE	78	8
65/35	9	12	39	15	6 4pt 6 3pt	39 MC 12 OE	81	8
80/20	9	12	48	15	6 4pt	48 MC 6 OE	72	8

English Composition

Our field-test events for English Composition will take place in Spring 2010 and Spring 2011. As you can see our plan includes a robust number of items to allow for attrition at data review. In addition, this plan would provide items to build all operational forms for the life of the contract, beginning with the first operational assessment in Fall (Dec.) 2011.

English Composition

Points Plan % OE & MC	Unique FT MC Items per FT Form	Common FT MC Items per Form	Unique FT OE Items per FT Form	Total FT Items per FT Form MC/OE	Total Points per FT Form	Total FT Forms
80/20	12	0	3 16 pt	12 MC 3 OE	60	45

U.S. History, World History, and Civics & Government

Our initial field test for U.S. History will take place in Spring of 2010. As you can see our plan includes a robust number of items to allow for attrition at data review. In addition, this plan would provide items to build the first 5 operational forms (4 core and 1 breach) from the initial field test for the administrations in Spring 2011, Summer 2011, and Fall 2011. The same plan will be used for World History and Civics & Government with an initial field test in Spring 2011 supplying items for the administrations in Fall (Dec.) 2011, Spring 2012, Summer 2012, and Fall (Dec.) 2012.

U.S. History, World History, and Civics & Government

Points Plan % MC & OE	# of Unique SS Item Sets per Module	Unique FT MC Items per FT Form	Unique Common FT MC Items per Form	Unique FT OE Items per FT Form	Total FT Items per FT Form MC/OE	Total Points per FT Form	Total FT Forms
50/50	4	30	15	9 4pt 6 2pt	45 MC 15 OE	78	8
65/35	4	39	15	6 4pt 6 3pt	54 MC 12 OE	81	8
80/20	4	48	15	6 4pt	63 MC 6 OE	72	8

VII.C.3. Embedded Field Test Passages/Items

DRC strongly believes in embedded field testing of items for future operational use. Embedding field test items within the operational assessment allows students to tryout items in as close to an actual administration experience as possible. This is in contrast to appending items at the end where fatigue and motivation factor heavily into student responses. With true embedding, as DRC is proposing, teachers, students, and administrators have no prior knowledge to which items are field test items, and therefore student responses are more reliable. The field test forms will be spiraled, ensuring student groups of essentially equal ability respond to each field test item. Please refer to *Subheading VII.B.2* for more information about DRC's embedded field test designs.

VII.C.4. Parallel Forms

To ensure that all test forms are equivalent DRC will start by matching the Test Blueprint and content standards specifications. Once valid content representation is accomplished, DRC will consider the Test Characteristic Curve (TCC) by form, ensuring that they match (coincident) for all forms. As discussed above in *Subheading VII.C.1*, DRC's psychometric staff will analyze the TIF and TCC curves with the goal that they match the target TIF and TCC as closely as possible given the constraints of the item pool and content specifications. This will provide the greatest assurance that the assessment will be consistent across administrations and any differences in student performance represent true changes.

Equating is a final check on form equivalency. Equating actually subsumes the calibration and scaling processes. DRC's equating processes were designed to help facilitate the accelerated reporting deadlines required with the RFP. DRC will also calibrate all field test items, placing the items on the item bank scale. This allows for forms construction to move forward considering all items in the bank and their associated difficulties. Rasch measurement methods will produce coincident TCCs if the same distribution of item difficulties is obtained.

DRC's test design discussed in the previous section provides for using 25–30% prior operational items, depending on content. This should foster scale continuity by enabling closer builds to the TIF and TCC targets. Additional information about equating can be found under *Subheading VII.J, Analysis of Assessment Data*.

VII.C.5. Breach Forms

DRC will construct two breach forms for each GCA course. The breach form will be constructed to mirror the core operational form, including number of items per reporting category. Each form will be parallel to the core operational form and will match the target psychometric guidelines.

DRC will follow the same forms construction process for the breach forms that we follow for the main GCA forms. The breach forms will be built and finalized

for printing. However, they will be printed, shipped, collected, and processed only if needed. Costs for printing and processing the breach form can be provided upon request.

VII.C.6. Plans for Construction of Multiple Test Forms

DRC will construct all test forms to comply with the test blueprints. The current test design includes core items and embedded field test items. Students' scores on the test are based on the core or common item set only.

Our recommendation is to make every attempt at providing for eight or more points per reporting category. Our design also provides for an adequate number of items field tested items each year for all subjects.

DRC will have:

- Coverage of the curriculum and ability ranges.
- Reliability of the score, including any based on sub-topics.
- Stability around the cut score.
 - With fewer points, there will be larger numbers of students who might be affected, up or down, by various rounding rules at the cut score.

DRC is prepared to work closely with PDE to develop specific item development plans and test designs as the program evolves and matures, ensuring a flexible and responsive approach to test development. The proposed test designs and item development plans are presented for planning and cost analysis purposes and reflect average constructs over the life of the contract.

Fall and Summer Retest

DRC will develop Summer and Fall Retest operational test forms each year for the GCAs. Retest operational test forms will be pre-equated to the GCA test scales and they will not include new embedded field test items. Please see *Subheading VII.J.* for more information about DRC's proposed linking design associated with the GCA forms.

VII.D. ASSESSMENT MATERIALS

DRC's Document Services business unit provides:

- In-house client-tailored publishing services.
- In-house customizable and exacting printing services.
- Consistently error-free products.

DRC takes great pride in the quality of testing materials produced on behalf of our department of education clients. We have extensive experience in producing assessment materials, including those for the PSSA for over 16 years, that are attractive and error free. Please see *Appendix 3* for samples of testing-related materials we have produced for our large-scale state assessment programs.

DRC will provide all GCA test booklets and answer documents, supporting materials, and ancillary materials in

appropriate formats and in sufficient quantities to districts and schools. The formats will maximize security and functionality. DRC recognizes that we will be responsible for producing, printing, and distributing GCA testing materials, including, but not limited to, test booklets, answer documents (including Braille and large-print versions), Directions for Administration Manuals, Assessment Coordinator's Handbook, Item and Scoring Samplers, Assessment Updates, and other necessary supporting materials, such as administration materials (school and district security checklists, packing lists, etc.). Mathematics formula and other reference sheets and general rubrics for open-ended items will be provided in test booklets and/or answer documents as appropriate.

Please note that DRC calculated all costs related to materials based on the quantities included on the Cost Submittal Worksheets provided by PDE. These quantities were sometimes different than those listed in the RFP or Appendices. Adjustments to materials quantities can be negotiated upon contract award. Our proposed materials specifications can be found in *Appendix 10*.

DRC project management staff will work closely with PDE to fulfill essential communication needs, develop accurate and engaging materials, and create aesthetically appealing test materials that appropriately address the quality requirements of PDE. Our proposed **GCA Component Lead, Ms. Karen Olsen**, will oversee the design, printing, and distribution of all testing-related materials, and in collaboration with other resource groups, will **ensure that PDE receives error-free products**.

Final draft materials will be provided to PDE for review and approval. Printing, publishing, and distribution will not begin until final approval has been provided by PDE. If desired, DRC will provide PDF files to PDE for posting on the GCA website as appropriate. DRC understands that any changes after sign-off will require PDE approval.

Using enrollment information and a 5% overage, all testing materials will be printed and distributed to Pennsylvania schools and districts in sufficient quantities. DRC will ensure that a sufficient number of copies are available for any last minute orders. Please refer to *Subheading VII.G.1., Customer Service Support*, for more information on DRC's customer service function and our user-friendly procedures for ordering additional materials.

GCA testing-related materials will be **produced through DRC's ISO 9001-certified Document Services business unit** that incorporates our Document/Graphic Design Group, which designs, edits, and typesets test materials, and our complete in-house Printing Department, which produces and prints scannable forms and other testing materials. All test booklets/answer documents will be developed in-house using our Adobe InDesign publishing system.

Mr. Tim Donahue, Senior Director of Operations, will provide direct management for the publications and printing operations for the program.

Mr. Donahue brings 34 years of experience in graphic arts management, with the majority of that time devoted to the production of printed products for large-scale data collection businesses.

Typesetting

DRC has complete, in-house publishing capabilities. Our Document/Graphic Design Group staff members are highly experienced desktop publishers, graphic designers, and editing professionals working with multiple publishing software systems. Our staff members are set apart by their expertise in development, proofreading, and production of educational assessment materials, as well as scannable forms design. This provides our assessment clients with unsurpassed expertise in design—the combined knowledge of both publishing and scannable forms design.

This incorporation of resources gives DRC a unique capability to customize our processes to address the requirements of each of our clients within restricted parameters and rigorous timeframes. DRC’s Document/Graphic Design Group routinely provides exemplary services to our state testing clients, including the detailed design, layout, and production of manuals, test booklets, and numerous ancillary materials, as well as the receipt, production, and transfer of electronic print files. **Our staff members are personally committed to developing accurate and engaging materials and creating aesthetically appealing test materials that meet all quality requirements.**

DRC and PDE will collaborate on updating and maintaining a program style guide to ensure consistent application of preferences and expectations across all program materials. The use of a style guide will ensure that detailed specifications for materials development are available to PDE and all DRC GCA Program team staff involved in materials development. It will also serve as the principal resource document to facilitate testing materials discussion between PDE and DRC, including our subcontractors as appropriate.

Ms. Debra Gartner, Document/Graphic Design Group Manager, will manage the critical activities associated with the composition and layout of all GCA Program assessment materials. She has nearly 18 years of experience with the development of Pennsylvania assessment materials, as well as assessment materials for other clients.

Ms. Gartner will ensure that produced documents have a look or style that is consistent with PDE stylistic preferences; she will further ensure that this style is consistent across all materials. Ms. Gartner and her staff have routinely provided exemplary services to PDE and other DRC clients, including the detailed design, layout, and production of manuals, test booklets, and numerous ancillary materials, as well as the receipt, production, and transfer of electronic print files.

Printing

DRC's Document Services business unit incorporates our complete in-house Printing Department, which produces scannable forms from typesetting to editing and printing. Because Document Services is under our direct control, DRC will be able to minimize our reliance on external resources that may have schedules that conflict with GCA Program deadlines. This means that we are able to allocate significant and superior in-house resources.



In addition to ISO 9001 certification, DRC has earned **Quality Level II status from the Government Printing Office (GPO)**, which is the second highest status that can be awarded (Quality Level I status is reserved for printers who produce bound books, four-color varnished promotional pieces, or other similar materials). To obtain this quality status, DRC underwent a rigorous review and inspection process, including work samples and process documentation.

The fine art of scannable document printing requires a mastery of the equipment that only years of training, experience, and dedication to the process can achieve. DRC stocks only those papers that provide the proper reflectance required by the leading OMR and imaging scanners. The same holds true for our inks; all have been tested and formulated to not only impart a crisp, clean look to the product for superior aesthetic appeal, but also to dropout in the proper measures for accurate data capture during the scanning process. Documents printed at DRC are printed to **exacting specifications to guarantee the highest possible data integrity for OMR, OCR, and Imaging machines**. DRC prints over 100 million documents a year. With each document, our skilled press operators hold our presses to incredible tolerances, exceeding even the strictest industry standards.

Specific areas of responsibility for Document Services staff involved in materials production will include: monitoring all materials production schedules to meet contract commitments; overseeing the production of scannable test materials; coordinating detailed printing and post-printing specifications outlining specific quality-control requirements for all materials; and conducting Print Reviews and Quality Checks. The printing staff will randomly pull documents to check for print quality throughout the print run.

DRC has successfully used these quality procedures for many state departments of education and is confident that our materials production protocols and standards will continue to meet all accuracy requirements.

Materials Production Quality Processes and Proof Sign-off



DRC understands that PDE requires error-free materials and that we have the ultimate responsibility for the quality and accuracy of the testing materials. DRC's Project Management, Test Development, and Document/Graphic Design Group personnel have extensive experience working with numerous state departments of education in the refinement and final presentation of test materials. These DRC departments know it is essential that all items in the test booklets and answer documents are in their intended and technically correct format. To this end, **DRC follows a meticulous set of internal quality standards** to ensure

high-quality printed products for its clients. DRC can assure PDE of its commitment to produce accurate materials. DRC will provide initial page proofs of manuals and other materials to PDE in either paper or electronic format, as requested.

Production Quality

All typesetting, desktop publishing, and printing activities associated with the production of GCA Program testing-related materials will be under our direct control. DRC's in-house Printing Department will print all scannable materials based on predetermined specifications and adhere to machine tolerance limits to ensure quality and accuracy. Any external printing and finishing vendors used by DRC to produce non-scannable materials will be pre-qualified using DRC's Subcontractor Qualification Procedure to ensure that all DRC subcontractors meet our stringent requirements for quality, accuracy, and on-time delivery of materials. DRC places the highest importance on this procedure and does not hesitate to deny work to subcontractors, or take existing work away from them, if our standards are not upheld. The selection of an external printing company will be predicated on that organization's adherence to internationally recognized quality assurance standards (i.e., ISO 9000). External vendor quality plans will be monitored by GCA project management staff and DRC's Quality Management Team. DRC will take full responsibility for the work performed by any vendor or subcontractor for the GCA Program.

DRC applies a meticulous set of internal and external quality standards to ensure high-quality printed products for our clients. DRC assures PDE of our commitment to produce 100-percent accurate, error-free materials. In order to meet these quality requirements, DRC staff will adhere to quality-control procedures throughout the production phase. The following outlines the quality processes we use to achieve error-free production:

Test Development Review and Quality Check—Test Development staff will confirm the consistency and quality of content and format across documents. They will work with Project Management to draft and review mock-ups of all forms to be used for this project and will work jointly with the

Document/Graphic Design Group to make certain that the instructions in the manuals match the test booklets, answer documents, and other materials. In addition, Test Development staff will review draft materials in conjunction with PDE staff to ensure error-free materials before the next step of production. They will send all proofs to PDE, allowing adequate time for review.

Publishing Review and Quality Check—The graphic designers, graphic artists, and word processing specialists in the Document/Graphic Design Group will work with Test Development and Project Management staff during the electronic publishing stage. They will ensure that all changes to publications are accurately incorporated. Typesetting style guides and editing conventions unique to the Pennsylvania assessments will be adhered to throughout the materials development process.

Before any material is submitted to PDE, DRC's Editing team will be responsible for coordinating word-for-word proofreading. A minimum of **three rounds** of independent quality checks (**word-for-word editing/proofreading**) will be conducted for each document. **At least two DRC Editors** will perform independent word-for-word reviews of materials. These specialists will provide suggestions for revisions to text and formatting where appropriate and will query potential issues to offset the possibility of misleading or technically problematic wording in directions and items. Edited copies will be returned to the electronic publishing staff for necessary corrections. Final proofs will be submitted to DRC's Editing team and to PDE for review and approval.

Materials Management Review and Quality Check—Test Development and Project Management staff will collaborate to review all test booklets and answer documents to confirm accuracy and continuity; they will query any test item that is suspect from a content standpoint. An external review will take place to verify the correct answers for selected-response items and to check completeness and accuracy of scoring rubrics for open-ended items. Test Development will work with PDE to revise scoring keys whenever necessary. These revisions will be verified and approved by DRC's Quality Assurance staff and PDE.

Information Systems Review and Quality Check—DRC's Quality Assurance, Information Systems, and Programming staff will work closely with Project Management staff to assist in scannable document verification. They will be responsible for certifying that all scannable documents are designed, developed, and printed within specified scanning requirements and tolerances. This technical review ensures error-free processing of scannable documents that will prevent delays in the data delivery and reporting schedule.

Printers Reviews and Quality Checks—DRC Project Management staff will work closely with internal and external print vendors during the print production phase. Once PDE has granted approval to print, DRC staff will be present during the first printing runs of each product and at various times during the collating, stitching, and trimming phases. DRC Test Development and Project Management personnel will check press proofs to make certain that the print quality meets

DRC standards (e.g., check for ink consistency, absence of smudge marks, precision of sequential numbering, collating and binding accuracy). If any errors are found, equipment will be stopped and the problem corrected. If reprinting is required, our printers ensure we have priority so that materials are completed by the scheduled deadlines so that packaging and distribution is not delayed. The vendor will pull every thousandth document to check print quality throughout the print run. Using enrollment information, all testing materials will be printed and distributed to Pennsylvania LEAs in sufficient quantities. Print quantities for testing-related materials will include a 10% overage to accommodate enrollment changes prior to test administration.

DRC's Project Management, Test Development, and Document/Graphic Design Group staffs have extensive experience working with numerous state departments of education, including PDE, in the development and production of printed test materials, including test booklets, answer documents, handbooks, manuals, forms, etc. These DRC departments know it is essential that all testing-related materials are produced in their intended and technically correct format. Project Management will provide PDE with written notification of any issues encountered during the materials development process and will consult with the appropriate DRC departments to suggest solutions to these issues.

VII.D.1. Assessment Booklets and Answer Documents

DRC will collaborate with PDE to develop accurate and aesthetically appealing test materials that appropriately address the rigorous quality requirements needed for the GCA Program. Test booklets (both non-scannable and consumable formats) and image-scannable answer documents will be developed in-house by our test development and document/graphic design specialists. DRC will ensure that the answer documents, test booklets, and other test materials are provided in appropriate formats to schools and districts. The formats will maximize security and functionality. DRC's **GCA Program Manager, Dr. Adisack Nhouyvanisvong**, and the **GCA Component Lead, Ms. Karen Olsen**, in collaboration with other resource groups, will ensure that PDE receives error-free products. The quantity of test booklets and answer documents produced will be based on enrollment data, as well as a 5% overage.

DRC proposes the following approach to test booklet and answer document format for the GCAs:

- Consumable (image-scannable) test booklets with integrated answer documents for English Composition. Due to the design and nature of this test, separating the response space from the multiple-choice items and prompts has no cost advantages and would only burden the administration of the test. The English Composition scannable test booklet will contain three test modules to accommodate student retests. This will allow students wishing to achieve a proficiency level in English Composition to retake only this module rather than the entire assessment.

- Non-scannable test booklets with separate, generic, image-scannable answer documents for all other subjects. A test booklet will be developed and printed for each course. Each course test booklet will contain three test modules to accommodate student retests. This will allow students wishing to achieve a proficiency level in a specific course to retake only the appropriate module rather than the entire assessment.

DRC has extensive experience producing both types of test booklets and answer documents for numerous state assessment clients. We welcome the opportunity to collaborate with PDE on test booklet and answer document design upon contract award.

Appendix 3 provides samples of test booklets and answer documents representing our proposed test booklet and answer document formats; please see our *Cost Submittal*, provided under separate cover, for costs associated with the proposed test booklet/answer document formats.

Test Booklets

Non-Scannable Test Booklets

For non-scannable test booklets, each course will have its own separate test booklet and a generic, image-scannable answer document. DRC has successfully designed and produced separate test booklets and answer documents throughout our PSSA contract.

All test booklets will be color-coded by course. The color scheme for the test booklets will be duplicated for the Administration Manual for each course. DRC will also develop Braille and large-print versions of the test booklets/answer documents (please see *Subheading VII.D.4* for more information). *Appendix 3* contains a sample non-scannable test booklet.

The test booklets will be customized to show the GCA name, the Pennsylvania logo, the form number, the course, and a place for student name on the covers. Each module will be divided into sections that can be administered either on the same day or on consecutive days.

DRC will ensure that all printing of GCA test booklets and other testing related materials adheres to strict quality-control procedures. Printed test booklets will be shrink-wrapped in packages of 17 with range sheets identifying the course and sequential security barcode numbers of the booklets enclosed. This packaging method and the use of Security Checklists (displaying the full range of secure materials within the shrink-wrapped packs) ensure that only essential LEA personnel and the student need to handle an individual booklet for testing.

Once the printing has been completed, DRC will provide PDE with printed copies of the test booklet, scannable answer document, and administration manual for each course.

Correlation between Test Booklets and Answer Documents

The test booklets and answer documents will mirror each other in terms of item numbering, section layout. Please see below for a discussion of our proposed procedures for developing and printing GCA image-scannable answer documents. The beginning and ending of test sections will be clearly identified in the test booklets and answer documents. DRC has successfully used such visual aids and prompts in Pennsylvania and in other states such as Alaska to help students navigate from test booklets to the appropriate sections of answer documents.

Test booklets and answer documents, including design, layout, use of graphics, and format of directions, will be developed according to Universal Design principles and procedures to minimize examinee confusion and ensure access by the broadest population of students. Additionally, clear, straightforward test instructions will also be provided in the administration manuals.

Consumable Test Booklets

English Composition will have its own image-scannable test booklet with an integrated response format. As with non-scannable test booklets, all consumable test booklets will be color-coded by course. The color scheme will be duplicated for the correlating Administration Manual for English Composition. DRC will also develop Braille and large-print versions of the English Composition consumable test booklet (please see *Subheading 4.C.6* for more information). *Appendix 3* contains a sample consumable test booklet.

Again, like the non-scannable test booklets, the English Composition consumable test booklets will be customized to show the GCA name, the Pennsylvania logo, the form number, and a place for student name on the cover. The consumable test booklet will be constructed to accommodate section breaks.

DRC will ensure that printing of the English Composition consumable test booklets adheres to strict quality-control procedures. Printed test booklets will be shrink-wrapped in packages of 17 with range sheets identifying the course and sequential security barcode numbers of the booklets enclosed. As described previously for non-scannable test booklets, the packaging method employed by DRC ensures that only essential LEA personnel and the student need to handle an individual booklet for testing.

Once the printing has been completed, DRC will provide PDE with printed copies of the English Composition consumable test booklet and Administration Manual.

Answer Documents

For all courses other than English Composition, separate customized, generic, image-scannable answer documents will be created in collaboration with PDE. Please see *Appendix 3* for a sample answer document.

In addition to recording student multiple-choice and open-ended responses, the answer documents will be used to collect test booklet form number and student

demographic data, as specified by PDE. Incorporated into each answer document will be a specially created section for adhering pre-printed student labels that contain student identification and demographic information. To accommodate those students who are enrolled after precoding has been completed, pre-printed district/school labels will be provided and student identification and demographic information will be hand coded into the demographic section. DRC will collaborate with PDE to determine design, layout, and location of the demographic sections.

The GCA answer documents will be designed, produced, and printed to fully support test administration, scanning, and scoring processes. The design and composition process for scannable answer documents is controlled by DRC's Document/Graphics Design Group in order to guarantee that all key departments, including scoring operations, have the opportunity to review and approve laser proofs and printer proofs.

DRC's Project Management Team will meet with PDE on an annual basis to determine if changes to the demographic sections of the answer documents are needed.

Test Booklet/Answer Document Security

DRC will adhere to strict quality-control procedures in the production of the test booklets and answer documents. We will generate a unique security barcode that will be printed on each test booklet and answer document. The barcode will ensure that each booklet and answer document can be unequivocally associated with only one record in a master database. Printed booklets and documents will be subjected to strict quality assurance inspections to ensure accuracy. The format and precision of the printed information will be closely examined by DRC's Software Quality Assurance Analysts to make certain the information is correct. DRC is confident that our use of barcoding technology will maintain an accurate account of all secure test booklets. The barcodes will allow us to maximize the capabilities of our Operations Materials Management System (Ops MMS), which is a proprietary and innovative system that utilizes barcode technology. Ops MMS provides an accurate, efficient, real-time method for tracking secure materials throughout every stage of administration, including packaging, distribution, and collection; materials receipt and check-in; processing, scanning, and editing; scoring; data validation and data conversion for reporting. The security barcodes are also used to create school security checklists, packing lists, and Missing Materials Reports.

VII.D.2. Mathematical Formula/Reference Sheets

DRC's Test Development Team, in collaboration with PDE, will develop mathematical formula sheets. The formula sheets will be revised, per PDE feedback each year of the program, if needed. DRC will make all PDE-approved revisions to the mathematical formula sheets and oversee the publication process for including the formula sheets in the appropriate printed materials (e.g., test booklet, Item Samplers). We will also supply mathematical formula reference sheets in a format

that can be posted on-line. Prior to publication, DRC will employ a series of quality-control steps designed to ensure that the mathematical formula sheets printed in materials (test booklets, Item Samplers) are free of error.

Should PDE request additional versions of formula sheets or other reference sheets, we would be pleased to work with PDE to develop these formula sheets or other reference sheets for inclusion in the appropriate testing materials.

VII.D.3. General Rubrics for Open-Ended Items

DRC's item and test development team, along with members of our performance assessment team, will develop new general rubrics for each course. We will make any revisions to these general rubrics should PDE request. The general rubrics serve as the foundation for the item-specific scoring guidelines for the open-ended items. The general rubrics for writing are mode-specific scoring guidelines and conventions scoring guidelines.

The general rubrics and the mode-specific and conventions general scoring guidelines for writing will be printed in the appropriate materials (i.e., test booklet, and/or answer documents) per PDE request. All general rubrics will also be printed in the Item Samplers. Prior to publication, DRC will employ a series of quality-control steps designed to ensure that the general rubrics printed in materials (test booklets, Item Samplers) are free of error.

VII.D.4. Large-Print and Braille Test Versions

Large-Print Versions

DRC has full, in-house capability to develop, produce, reformat, and print large-print test materials. We have extensive experience providing large-print materials for many state assessment clients, including Alaska, Louisiana, Idaho, and South Carolina.

DRC's Document/Graphics Design Group and Printing Department will provide large-print materials formatted to meet the needs of the GCA Program. We will work with PDE during typesetting to ensure all specifications for the large-print versions are met. The large-print materials will be provided to PDE for review and approval prior to production. Supplemental instructions regarding transferring of student responses to the test booklets/answer documents will be provided to test administrators. Using quantity requirements provided through the enrollment process and/or by PDE, DRC will ensure sufficient quantities of large-print versions of test materials are available for each administration.

Braille Versions

Each year, DRC will work with staff from American Printing House for the Blind (APH) to produce a Braille test booklet for each GCA test. APH will ensure that all tests are modified correctly and that they are accurate. APH has produced Braille versions of tests and educational materials for numerous publishers and testing organizations. DRC has used their services for other assessment programs

for several years with excellent results. Both contracted and uncontracted Braille will be available, and codes accepted by the Braille Authority of North America (BANA) will be followed.

Supplemental instructions regarding transferring of student responses to the test booklets/answer documents will be provided to test administrators. Braille versions will be ordered during the enrollment process, and DRC will ensure that a sufficient number of copies are available for any last minute orders.

VII.D.5. State Approved Accommodations

DRC is keenly aware of federal mandates that require assessments to be accessible to all students regardless of ability. DRC has worked with the National Center on Education Outcomes (NCEO) on many assessment projects, including the design and development of alternate assessments for students with disabilities and English Language Learners, as well as designing and facilitating standard settings for alternate assessments. We stay abreast of current legislation, technologies, and methodologies that impact universally designed assessments and retain staff members that are knowledgeable, trained, and experienced in developing and administering assessments that are accessible to all students. DRC is ready to assist PDE with the development and administration of GCA tests that are universally designed, provide for test accommodations/modifications, and meet state and federal legislation requirements.

VII.E. ANCILLARY MATERIALS

DRC will work collaboratively with PDE to develop ancillary materials that effectively communicate consistent messages to the intended users. DRC will ensure that each proof is free of typographical and formatting errors before submission to PDE for review and approval. Any changes after PDE sign-off will require PDE approval.

DRC successfully produces similar types of materials for many of our other state assessment clients. We take pride in producing clear, easy-to-understand, aesthetically pleasing, and error-free informational and interpretive materials. Please see *Subheadings VII.D., Assessment Materials*, for additional information on DRC's materials production procedures.

VII.E.1. Assessment Update Bulletins

DRC will produce and distribute up to six editions of the Assessment Updates annually. The multi-page (1–4 pages) Assessment Update newsletter will be distributed to all school and LEA Assessment Coordinators and other school personnel, as directed by PDE. The Updates will be available electronically, as well as hard copies that are shipped to schools/district. The Assessment Updates will be co-authored by DRC's Project Management team and PDE.

The Assessment Updates will contain useful and relevant assessment information. Some possible topics for Assessment Update newsletters include:

- Enrollment verification information
- Upcoming key dates
- Training/staff development announcements/applications
- Field test announcements/applications

VII.E.2. Item and Scoring Samplers

DRC will develop Item and Scoring Samplers each year. The Item Samplers will include released items for each GCA assessment. They will also include answer keys for the multiple-choice items, along with distractor analysis for the multiple-choice answer choices and scoring guidelines for open-ended items and writing prompts.

In addition, the samplers will include authentic exemplars of students' written responses to open-ended items or writing prompts. DRC has a staff of experienced handscoring personnel well versed in selecting sample student responses for state release. DRC is aware that annotations for a public audience must be written differently than training annotations. In conjunction with our Test Development staff, we have provided this service for the PSSA for the past several years, as well as for many of our other clients, including the states of Kentucky, Louisiana, North Carolina, and South Carolina.

DRC will prepare the Item and Scoring Samplers for electronic posting on PDE's website. Prior to publication, DRC will employ a series of quality-control steps designed to ensure that Item Samplers are free of errors. Please see *Appendix 3* for sample Item and Scoring Samplers.

The Item and Scoring Samplers will include released items after the first operational administration of each assessment. The Item and Scoring Samplers must include sample items for each assessment anchor, along with the performance statistics from the prior operational assessment.

VII.E.3. Administration Manuals and Assessment Coordinator's Handbook

Administration Manuals

We understand that testing is an enormous undertaking for all school staff at a very busy time of year, so we constantly strive to make test administration directions more useful and easier to understand. DRC has extensive background and experience in writing and editing effective manuals, including those used for the PSSA. DRC will work with PDE to design Administration Manuals that are clear, concise, and user-friendly, making it easier for test administrators to be successful. The test administration manuals will contain easy-to-follow directions written in narrative format to be read directly to the students.

DRC anticipates developing an Administration Manual for each course. To streamline the process for busy test administrators, each course Administration

Manual will encompass all three modules. The manuals will include information pertaining to the handling and security of the test booklets and answer documents and general information about how to administer the GCAs, as well as specific test instructions for each test. DRC staff will work closely with PDE to establish and update policies and procedures for the administration of the GCAs.

The manuals will be mocked up, typeset, and submitted using similar developmental and proofreading steps as other GCA testing materials. During the development process and prior to printing, a four-way match between the test booklets, answer documents (or integrated test booklets), Administration Manuals, and Assessment Coordinator's Handbooks will be performed by DRC Project Management and Test Development staff to ensure accuracy of all instructions. Manual proofs will be free of typographical and formatting errors before they are submitted to PDE for review.

As with all materials, PDE will review the content and format of the manuals and will have final approval. After incorporating any PDE edits and revisions, DRC will print and package the Administration Manuals. Hardcopies of the manuals will be distributed to districts in appropriate quantities based on enrollment information at a rate of one manual for every 15 assessment booklets; they will be shipped to arrive no later than four weeks before the testing window. As with the Assessment Coordinator's Handbook, the Administration Manuals will also be provided in PDF format for use online. A sample Administration Manual is included in *Appendix 3*.

Assessment Coordinator's Handbook

DRC will work collaboratively with PDE to develop and update the Assessment Coordinator's Handbook so that it continues to effectively communicate consistent information. Throughout the Assessment Coordinator's Handbook, graphic illustrations will be used where appropriate to clarify GCA procedures. The handbooks will be made as user-friendly as possible. DRC successfully produces similar handbooks for many of other assessment programs, including the PSSA. We take pride in producing clear, easy-to-understand, aesthetically pleasing, and error-free assessment coordinator handbooks.

DRC will collaborate with PDE to develop an Assessment Coordinator's Handbook that provides the following information:

- Description of the GCAs.
- Responsibilities of the district assessment coordinators/building assessment coordinators, test administrators, and monitors.
- Training of assessment coordinators.
- Procedures for receipt, distribution, collection, and return of materials to DRC for processing and scoring.
- Preparation for the assessment.

- Student inclusion rules.
- Procedures for maintaining security.
- Logistical information for appropriate administration of the tests, including testing guidelines and scheduling.
- Contact information to directly connect with DRC's live Pennsylvania Customer Service team.

We propose that one Assessment Coordinator's Handbook be produced to include all courses. The Assessment Coordinator's Handbook will be mocked up, typeset, and reviewed using similar development and proofreading steps as the test booklets/answer documents. PDE will have the opportunity to revise and approve the Handbook prior to printing. After incorporating any PDE edits and revisions, DRC will print and distribute the handbooks in appropriate quantities using enrollment information and or quantities required by PDE. The Handbooks will be shipped in the first shipment with the non-secure testing materials at least four weeks prior to the testing window. Handbooks will also be provided electronically in downloadable PDF format for use online. *Appendix 3* contains a sample Assessment Coordinator's Handbook.

VII.E.4. GCA Guide for Students and Families and GCA Parent and Family Website

GCA Guide for Students and Families

Providing parents and students with information about the GCA Program is a critical aspect of this assessment program. Providing parents with information directly related to their child's academic work can have a positive effect on student learning. Therefore, providing parents with information that they are comfortable with, and which is engaging, holds even greater promise for positively affecting student performance.

DRC will design, produce, and disseminate a GCA Guide to help students and parents become familiar with the GCA Program, including the purpose of the GCAs, the content that will be assessed, and the types of items that will be included in the assessment.

DRC will ensure that each proof is free of typographical and format errors before submission to PDE for review. The finalized GCA Guides will be provided electronically in PDF format for posting on PDE's website and/or the GCA Parent and Family Website.

GCA Parent and Family Website

DRC welcomes the opportunity to again collaborate with PDE to design and develop a parent website—DRC designed and developed the original PSSA Parent Website.

The GCA Parent and Family Website will be designed to provide helpful suggestions and useful links for Pennsylvania parents. The site will be a tool that provides parents with essential information to assist them in playing an active role in their child's educational success. As with any website, content is a critical element. The website will provide current information and significant content for parents to use when interpreting their child's test results and will enhance their overall understanding of the GCA Program. Upon contract award, DRC will meet with PDE to determine the design, format, and content of the GCA Parent and Family Website.

After the website has been launched, DRC will implement customized updates and revisions based on the needs of PDE and parents. DRC will work closely with the PDE to ensure that the updates and revisions mirror the original design requirements. At the same time, DRC will work in conjunction with the PDE to provide the content updates that are essential in providing parents with the tools and information necessary to develop their children's education. DRC will revise the website, as requested by the PDE, when changes to the assessment programs occur. and as student reports are distributed.

DRC will continually monitor the performance of the website to ensure that any hardware or software issue that may arise is resolved in a timely manner to have the least amount of impact on parents. The website, linked from the PDE homepage, will reside on DRC's server and will allow DRC to facilitate any software updates and revisions. DRC will comply with Commonwealth policies and restrictions regarding the Internet and Intranets as well as adhere to web accessibility guidelines.

Software Quality Assurance Testing

DRC's Quality Assurance Analysts will ensure that each website page, link, and image displays properly, follows Graphical User Interface (GUI) standards, and functions as designed. Special quality checks will be executed to ensure content is correct and accurate. All website changes and modifications will be tested on a dedicated test server prior to being released into the production environment. Additionally, the website will be tested on various computer platforms, using multiple browsers, and numerous browser versions to ensure compatibility with the majority of the general public. Once moved to the production server, the Quality Assurance Analysts will again verify that the website is accurate and is ready for public access.

VII.F. MATERIALS DELIVERY

VII.F.1. Produce and Distribute Assessment Materials

VII.F.1.a. Determination of Materials Quantities

DRC will use the information collected through our online Enrollment Verification System described below to establish a master database of enrollment data and LEA and school/testing site addresses and current contact (e.g., assessment coordinator, PIMS administrator, superintendent), which will be shared with PDE for approval. From this database, DRC will determine final material print quantities and produce control forms, such as packing lists and security checklists. DRC will apply a 5% overage for printed test materials. We will work with PDE to adjust overage percentages as needed.

VII.F.1.b. Enrollment Procedures

Each year, DRC will work with PDE and LEAs to verify student enrollment counts by course for each school participating in the assessments and provide LEAs an opportunity to update their contact and shipping information. LEAs will also be offered the opportunity to designate whether they will be taking the test online, via paper, or both.

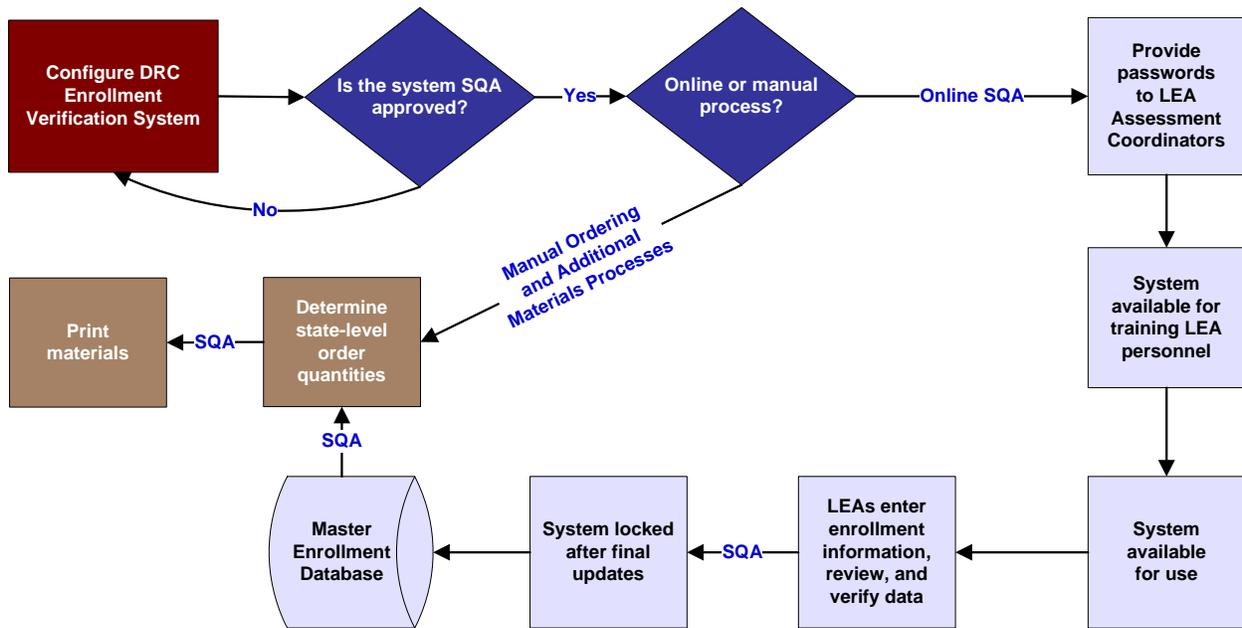
To accomplish this, DRC proposes using our online **Enrollment Verification System** to collect and update school enrollment information and testing mode preference (online or paper or both). This **web-based solution has been successfully utilized for the PSSA** and by our other large-scale assessment clients and is known for its convenience, accessibility, and ease of use. The online Enrollment Verification System will be accessible through PDE's Ed Hub.

DRC will utilize PDE databases, such as EdNA or PIMS, as needed to ensure accurate information. LEA Assessment Coordinators will access the system to confirm or modify the quantities for their specific schools/testing sites. DRC will implement integrated quality checks to ensure final enrollment counts are closely aligned with the projected enrollment counts to help guard against inflated material quantities.

DRC's proposed enrollment process is outlined in the figure that follows. We are confident that our proven process, coupled with our flexible online system, will continue to meet PDE's requirements and will provide LEAs with a **straightforward, user-friendly enrollment process**.



DRC Enrollment Process



DRC Enrollment Process

This system provides PDE and LEA personnel with many advantages:

- Access a secure website that is password protected for each LEA.
- Pre-populated system with quantities from previous administrations.
- Review a list of materials by school/grade available for order for each test administration.
- Input or modify order quantities for regular and accommodated testing materials.
- Submit all orders electronically.
- Receive confirmation notices of enrollment quantities.
- Move summary level orders accessible by PDE.

DRC will work with PDE to establish a schedule for enrollment collection. DRC will email or fax a memo to LEA Assessment Coordinators, informing them of the enrollment timeline and providing general instructions for accessing the online system. In addition, DRC Customer Service Representatives will be available to answer any questions regarding the online Enrollment Verification System, or the enrollment process in general, prior to and during the enrollment window.

LEA Assessment Coordinators will use the Enrollment Verification System to easily submit, review, and update material quantities for each school/testing site participating in the assessment. The following figure displays a sample of the system's enrollment entry screen.

DATA RECOGNITION
DRC
CORPORATION

Home Enrollments Maintenance Addresses Summary User Manual Log Out

District: 999999999
PENNSYLVANIA DISTRICT
School: 999999999
PENNSYLVANIA SCHOOL

School: PENNSYLVANIA SCHOOL

Subject: Reading, Mathematics, Writing, and Science

Grades	School Enrollments	PSSA - M	Spanish Translation	Large Print	Braille
Grade 3	<input type="text" value="0"/>				
Grade 4	<input type="text" value="0"/>				
Grade 5	<input type="text" value="0"/>				
Grade 6	<input type="text" value="0"/>				
Grade 7	<input type="text" value="0"/>				
Grade 8	<input type="text" value="0"/>				
Grade 11	<input type="text" value="0"/>				

Click Submit to save or confirm enrollments.

Copyright © 2005 Data Recognition Corporation. Patents Pending.

**Online Enrollment Verification System—
Enrollment Entry Screen**

The system will also allow updates to school names, LEA Assessment Coordinator names, addresses, email addresses, telephone numbers, and fax numbers. See the following figure for a sample of the system's address maintenance screen.

DATA RECOGNITION
DRC
CORPORATION

Home Enrollments Maintenance **Addresses** Summary User Manual Log Out

District: 999999999
PENNSYLVANIA DISTRICT
School: 999999999
PENNSYLVANIA SCHOOL

District: PENNSYLVANIA DISTRICT

Address Type: Assessment Coordinator Shipping

* Denotes Required Field

Assessment Coordinator Shipping - Primary Information

Contact Name * TEST TESTER
Contact Title TESTING COORDINAT
Phone * (717) 787 - 0000 Ext. 123456
Fax (717) 787 - 0000
E-Mail

Address Information
P.O. Box addresses cannot be accepted for shipping addresses.

District PENNSYLVANIA DISTRICT
Address 1 * 1234 PENNS AVEN
Address 2
City * HARRISBURG
State * PA
Zip * 17109

Submit Cancel

Copyright © 2005 Data Recognition Corporation. Patents Pending.

Online Enrollment Verification System— Address Maintenance Screen

VII.F.2. Collection of Student Demographic and Program Information

Student Data Collection

DRC is accustomed to retrieving data from student information systems and maintaining the accuracy of state-assigned student identification numbers. Many of our state assessment clients use such state student identification systems. Student identification and demographic information, LEA and school/testing site codes, birthdates, gender, etc., can be easily loaded into DRC's master database system.

For the 2008 PSSA administrations, DRC successfully implemented processes to receive student demographic data from the Pennsylvania Information Management System (PIMS). Building on our four-year history of successfully collecting 92% of the public students' data at the district level with accuracy, DRC worked closely with PDE and their partners to define the required data elements and to determine schedules and procedures to ensure a successful transition into data collection through PIMS. With an understanding of the need for data integrity and the importance of longitudinal reporting and tracking, DRC played a significant role in assisting PDE with the assignment of state IDs (PAsecureID) to all students.

We understand that all student identification, demographic, and any score attribution information required for reporting will be provided via PIMS. DRC will incorporate the PIMS information into our system and process the data so that a master precode database is established. This data will be used for the material production and precode label production. Detailed standards during the data transfer process will be followed and quality inspections will be performed by DRC's Software Quality Assurance Analysts to ensure the data is transferred accurately. DRC will perform detailed validation on the data files received. Data issues that affect the accuracy of reporting a student's results will be presented to PDE for resolution.

Prior to final reporting, if required, DRC will accept a second file transfer from PDE/PIMS that contains updated precode data for the testing students. The updated data will be incorporated into our master database for use during student and summary reporting.

DRC is assuming that all student information needed for scoring will be coming from PIMS (e.g., attribution information) since PDE did not request online precode systems or correction systems in the RFP or the RFP Questions and Answers. If either or both of these online systems are desired by PDE, they can be negotiated upon contract award.

Procedures for Interfacing with PIMS

The exchange of data between entities is a critical and essential component in the continued success of an assessment program like the GCAs. To support this process, DRC proposes using our data exchange procedures to ensure that all data files are accurately and securely transferred to DRC from PDE/PIMS. We recognize the importance of this function and have embedded quality checks throughout. DRC will work with PDE to confirm these procedures and will modify the process as appropriate.

Please refer to *Subheading VII.K., Reports and Data Files*, for more information on DRC's data management procedures.

Security of Precode Information

DRC will incorporate rigorous quality assurance activities throughout the process to ensure the highest level of data quality, integrity, and security. All precode data will be accurately stored in a secure database environment. In our computing environment, DRC utilizes security controls that relate to our hardware, data, and network. DRC manages multiple terabytes of client data; therefore, security is an inherent, inextricable, and indispensable component of our system. DRC has extensive experience in designing systems for our clients that have built-in audit trails. All systems developed and used for the GCA Program will include audit trails.

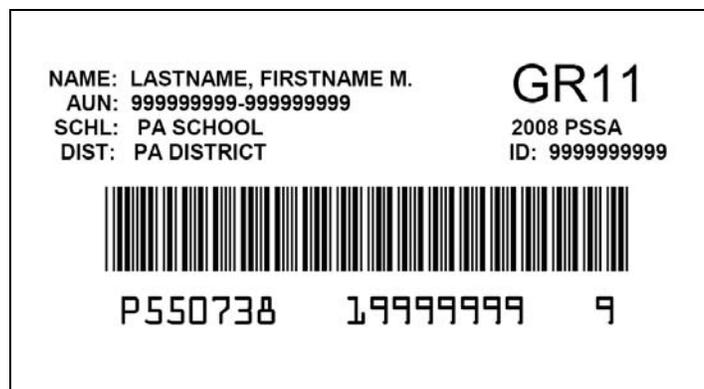
DRC will ensure that all student data remains confidential and secure. Individual student data records will not be distributed unless specifically requested by PDE in writing. DRC is highly aware of the confidentiality of student information. Each DRC employee is keenly aware of the critical need to maintain student data confidentiality and security.

Please refer to *Subheading IV.C.15., Security of Test Materials and Results*, for more information on DRC's test security features and procedures.

VII.F.3. Pre-Code Labels and Pre-Code Student Information

For each participating student, DRC will produce and distribute student precode labels based on the data received from PIMS. The information presented on the label will include a unique DRC precode number and barcode. The information can also include student last and first name, PAsecureID, school/testing site name and number, LEA name and number, course, and other administration information required by PDE.

The precode barcode will ensure that each document returned to DRC can be unequivocally associated with only one record in our master precode database. The following figure displays an example of the type of precode information printed onto student labels.



Sample Student Precode Label

Upon receipt of the used student test booklets/answer documents at DRC, the barcode number on the student labels affixed to the test booklets/answer documents will be scanned and validated against the master precode database. The barcode corresponds to the precode number and is human- and machine-readable.

Printing of the precode labels will take place at DRC's in-house printing facility. Requirements and printed materials will be subjected to strict quality assurance inspections to ensure accuracy. The format and precision of the printed information will be closely examined by Software Quality Assurance Analysts to make certain the information is correct.

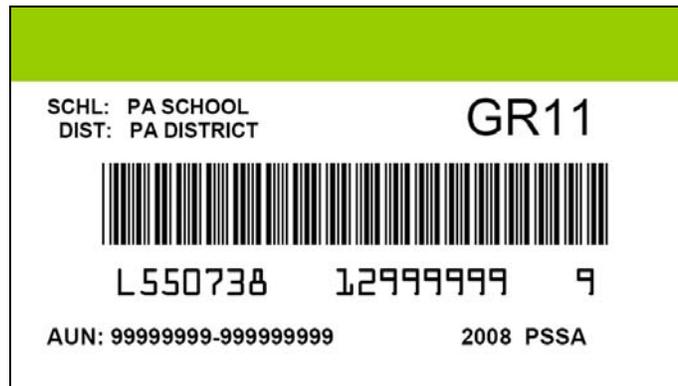
Student Mobility

Our system and processes are uniquely designed to address student mobility. Students who are not included in the master precode database, such as new or transfer students, can be hand-coded on their answer document, using procedures provided in the handbook and manuals.

To accommodate students not included in the master precode database at the time of test administration, DRC will provide LEAs and schools/testing sites with additional test booklets/answer documents. For the purpose of identifying these students, DRC will provide district/school labels for each school. These labels will then be affixed to test booklets/answer documents for students who do not have precode labels. Each label will also contain a unique barcode/number that will be used to associate the test booklet/answer document with a specific district and school. Student identification and demographic information will then be hand-coded into the demographic section of the test booklet/answer document. For these students, identification and demographic information will be incorporated into DRC's master database at the time of scanning. Hand-coded information is verified through our editing process to ensure 100% accuracy (please refer to *Subheading VII.I. Scanning/Imaging and Scoring*). Each student included in the master precode database or from hand-coded test booklets/answer documents will be accounted for and included in the scored student file, ensuring a high rate of accuracy and accountability.

District/school labels will be collated with the student precode labels in district and school order; LEA Assessment Coordinators will be able to order additional district/school labels and test booklets/answer documents through DRC's Pennsylvania Customer Service function (please refer to *Subheading VII.G.I., Customer Service Support*). Please see *Subheading VII.I* for more information on DRC's processing and scanning processes, software programs, and quality-control procedures.

A sample district/school label is depicted in the following figure.



Sample District/School Label

DRC has many options to offer for updating and maintaining student demographic data. We look forward to customizing a solution that best meets PDE’s needs and welcome the opportunity to further discuss other precode options.

VII.F.4. Assembly, Packaging, and Shipping of Assessment Materials

DRC has fulfilled packaging and shipping requirements for numerous large-scale assessment programs, including Alabama, Alaska, Louisiana, Ohio, Oklahoma, Pennsylvania, and South Carolina. Our **ISO 9001:2000-certified distribution process** underscores the importance of quality standards; we take all necessary precautions to ensure accurate packaging and timely delivery of all materials.

DRC’s proprietary materials management system, Ops MMS, ensures:

- Accurate, efficient packaging.
- Secure tracking of barcoded documents in all phases.
- 100% accounting of all returned secure materials.

DRC’s **Director of Materials Operations, Mr. Doug Miller**, will coordinate the packaging of GCA testing materials by Operations staff, who will use DRC’s proprietary Operations Materials Management System (Ops MMS). This proprietary system provides an accurate and efficient method for packaging materials. Systematic quality controls facilitate the tracking of secure materials throughout the packaging and distribution phase. Using scanners to “scan out” order-specific materials, Ops MMS provides the flexibility to package secure documents already

pre-assigned at a site-specific level, or to use barcode scanning to assign secure ranges at the point of packaging. All requirements provided by site-specific packing lists must be satisfied by these scans in order to complete the packaging process. These systematic controls ensure that the accurate quantity, material type, and security range are pulled and packaged for the appropriate site.

Through Ops MMS, DRC can view data on the items scanned into any box and compare this data to the physical box contents. Random boxes of packaged material are pulled, opened, and verified for accuracy against the contents listed in the system. All data generated during packaging will be made available to our

Project Management team, providing a tool for monitoring shipments and satisfying client concerns.

All test materials, including a packing list, an acknowledgement of delivery form, and other shipping materials, will be packaged by school and shipped to the district or non-district testing sites, such as charter schools, full-time Career and Technology Centers, alternative education programs, approved private schools, and Intermediate Units. To maintain consistency with the PSSA, DRC proposes that for large districts with ten or more buildings, testing materials will be shipped directly to the schools. All boxes will be labeled with the test coordinator's name, address, and "ATTENTION: SCHOOL TEST COORDINATOR."

Secure test materials will be spiraled to ensure equal forms distribution at the classroom level and shrinkwrapped in packs of 17. Distribution quantities for secure testing materials will include a 5% overage based on quantity information obtained through the enrollment process and/or PDE's PIMS system. DRC will ensure that a sufficient number of copies are available at our facilities for any last minute orders.

For most shipments, DRC uses a standard 11" box with a minimum of 275 pounds bursting weight. Box size will vary to accommodate specific shipping demands, but will be small enough for easy handling. All fully packed boxes will weigh less than 30 pounds. Each shipping box will be pre-printed with DRC's return address and affixed with a large, brightly colored label stating "PENNSYLVANIA GCA MATERIALS—OPEN AND INVENTORY IMMEDIATELY. ITEMS ARE SECURE." We encourage districts to re-use these boxes for the return shipment, but we will send additional boxes to districts who request them. Return labels containing school and district information will be included to simplify the return process.

Each site will receive at least two shipments. The first shipment will contain non-secure materials, including the Assessment Coordinator's Handbook, Directions for Administration Manuals, and School and District Assessment Coordinator Packets, with copies of all forms and special instructions, and will be shipped to arrive at each site no later than four weeks prior to the start of the testing window. The second shipment will contain all secure testing materials, including test booklets and answer documents, as well as precode labels; these materials will be shipped to arrive no later than two weeks before the testing window.

Assessment coordinators will be required to submit a Materials Receipt Notice (acknowledgement of delivery) and fax or mail it to DRC. The forms will be logged in based on a list of participating LEAs/testing sites. DRC will contact coordinators to confirm delivery for any instances in which acknowledgements of delivery were not received within seven days of shipping.

Assembly and Packaging Accuracy

DRC will ensure that all assessment materials are assembled correctly prior to shipping using the following approach:

- Based on requirements and specifications stipulated in the contract and gathered through discussions with PDE, detailed Scope of Work Agreements (SOWAs) will be established by the GCA Program Manager, working in conjunction with experienced Project Management and Operations staff. These SOWAs will ensure that all staff understand and adhere to materials assembly and distribution requirements. The SOWAs will be available for PDE review at each step of the process.
- The GCA Project Management team will conduct a “walk-through” prior to each shipment to ensure that all assembly and distribution procedures are followed precisely.
- DRC Operations staff will use our innovative and proprietary automated Operations Materials Management System (Ops MMS) to assign items to the appropriate site for shipment. This system uses barcode technology to provide an automated quality check between items requested for a site and items being shipped to a site based on the following:
 - Project
 - Site (school, district, or other testing location)
 - Subject
 - Form (if necessary)
 - Material type
 - Quantity
- Should any discrepancies occur between the materials being requested for a site and the materials being packaged for a site, the discrepancy will be resolved before the order is completed and shipped.
- A shipment box manifest will be produced for and placed in each box. DRC Operations staff will double check all box contents with the box manifest prior to the box being sealed for shipment.
- DRC Operations staff will perform lot acceptance sampling on every shipment. Two to three LEAs will be selected from each page of the shipping roster and examined for correct and complete packaging and labeling. This sampling represents a minimum of 10% of all shipping sites.

Shipping Accuracy

DRC will ensure that materials are shipped to the proper locations using the following approach:

- DRC's Director of Materials Operations and Logistics Manager will continuously monitor packaging and coordinate shipping processes.
- At the time of shipping, our proprietary materials management system, Ops MMS, interacts with DRC's Education Project Information Center (EPIC) to utilize accurate, current shipping information. This allows for the incorporation of real-time shipping information updates and eliminates the potential for error associated with maintaining multiple shipping information databases. Shipping addresses and local testing coordinator contact information for each participating school and district are stored and maintained in EPIC. EPIC system-checks allow for detection of potential errors, such as duplicate entries and misspellings; errors are flagged for resolution by the Customer Service team. EPIC also verifies that shipping addresses are valid. Please see *Subheading VII.G.1., Customer Service Support*, for more information regarding EPIC and our Customer Service function.
- All shipping labels will be quality checked to prevent materials from being sent to the wrong location. Site labels on each box will be compared to the shipping address label and matched for accuracy.
- United Parcel Service (UPS) and **two local Pennsylvania freight carriers, Diamond Transportation based in Philadelphia and Advanced Shipping Technologies (AST) based in Pittsburgh**, will be used to ship materials to LEAs and schools. These carrier methods will allow cost-effective, online traceable, and timely distribution of materials.
- Ops MMS will integrate with shipper systems, allowing the Program Manager and the Logistics Manager to track materials from the point of shipment from DRC's warehouse facility to receipt at the LEA testing site.
- Assessment coordinators will be required to submit a Materials Receipt Notice (acknowledgement of delivery) and fax or mail it to DRC. The forms will be logged in based on a list of participating LEAs/testing sites. DRC will contact coordinators to confirm delivery for any instances in which acknowledgements of delivery were not received within seven days of shipping.

DRC Packaging and Shipping Quality Procedures

- **Detailed instructions**—Based on contract requirements and specifications, detailed Scope of Work Agreements (SOWAs) will be established by the GCAs Project Management Team working in conjunction with our Operations staff. The SOWAs will be available for PDE review at each step of the process.
- **Walk-throughs**—The Project Management team will conduct a walkthrough of the assembly process prior to each shipment to check that all procedures are precisely followed.
- **Ongoing monitoring**—The Director of Materials Operations and the Logistics Manager will monitor the materials assembly area and report any irregularities to Project Management.
- **Secondary checks**—Our Operations staff will perform secondary checks on all packing lists before boxes are sealed for shipping.
- **Address accuracy**—All district and school shipping labels will be quality checked to prevent materials from going to the incorrect location. Site labels on each box will be compared to the shipping address label and matched for accuracy.

VII.F.5. Packing Lists and Delivery Notification

DRC acknowledges that a letter outlining the mode and projected date of delivery, along with a copy of the packing list(s), must be sent under separate cover to each site.

VII.F.6. Materials Delivery

DRC will ship all testing materials directly to LEA offices or, for large districts with ten or more buildings, directly to the schools. Delivery of materials will be scheduled during regular weekday school hours, 9:00 a.m. to 3:00 p.m. Eastern Time, or by appointment with school officials. DRC will send email notifications to test coordinators when materials are shipped. All shipments will be designated as “inside delivery required.” Signatures of receipt will provide proof of delivery and allow DRC and LEAs to track all shipments via UPS QuantumView™ and Diamond and AST’s websites. PDE will also have access to these online tracking systems, if desired.

When LEAs receive their shipments of secure materials, they will be asked to complete a Materials Receipt Notice (acknowledgement of delivery) and fax or mail it to DRC. The forms will be logged in based on a list of participating LEAs/testing sites. If the forms are not received within 7 days after the shipment was made, DRC will contact LEAs to verify receipt of materials. Upon request, DRC will provide PDE with a proof of delivery report, which will include a summary of undelivered materials.

VII.F.7. Customer Service for Materials Distribution

LEA, school, and PDE personnel will have the complete support of DRC's courteous, professional, "live" Customer Service team to assist with package tracking, resolution of delivery issues, and/or ordering additional materials. Our customer service function is organized such that only staff trained in the GCA Program will respond to calls. DRC's toll-free number and email address will be prominently listed in the Directions for Administration Manual, the Assessment Coordinator's Handbook, and other correspondence with schools and LEAs. Please see *Subheading VII.G.1., Customer Service Support*, for more information on DRC's Pennsylvania Customer Service function and process.

Additional Materials

DRC maintains its superior customer service reputation by making sure orders for additional materials, including additional accommodated materials, are shipped promptly. Should a school's enrollment change between the submission of its enrollment and the time of testing, the Assessment Coordinator may request additional testing materials. To order additional testing materials, including any accommodated materials, the Assessment Coordinator need only complete the additional materials request form provided in the Handbook and fax it to DRC's Pennsylvania Customer Service Team. Assessment Coordinators may also call a Customer Service Representative to order materials over the phone or submit an order via email. DRC's toll-free number and email address will be listed prominently throughout the administration materials.

DRC's Customer Service Team uses DRC's Order Fulfillment System to instruct our Logistics Department to send the requested materials to the testing site. This proprietary system allows flexibility of delivery locations and methods and is linked to Ops MMS for overall tracking of secure materials.

Requests for additional materials received after 3:00 p.m. Central Time will be processed on the following business day. DRC will ship orders via secure, traceable ground delivery, unless 2-day or overnight is warranted to ensure materials arrive before testing. Records of orders for additional materials will be maintained by the GCA Project Management Team along with the records of original shipments and will be made available to PDE for review, if desired. Date of shipment, addressee, document or material shipped, shipping weight, and method of shipment will be indicated so that proper documentation will be available should questions arise regarding the shipment and/or receipt of materials. Please see *Subheading VII.G.1., Customer Service Support*, below for more information on DRC's Pennsylvania Customer Service function and procedures.

VII.F.8. Report of Packaging and Distribution Activities

Immediately following packaging, DRC will provide PDE with a packaging and distribution report, which will list each distribution site and will include for each site the carrier used, the shipping date, and the number of pieces shipped.

VII.G. ASSESSMENT ADMINISTRATION SUPPORTS

VII.G.1. Customer Service Support

DRC takes great pride in our record of customer satisfaction, and this attitude will be apparent as we continue to respond to both the needs and requests of PDE and Pennsylvania School and District Assessment Coordinators. In order to ensure ongoing customer satisfaction, DRC will provide customer service support consisting of GCA-experienced, professional, trained, and responsive personnel who understand all dimensions of testing programs and who are in a position to act decisively to resolve project challenges. Our proposed **GCA Component Lead, Ms. Karen Olsen**, will oversee the customer service for the GCAs.

DRC's Pennsylvania Customer Service Representatives are:

- Client-focused, experienced, courteous, and responsive.
- Knowledgeable of all dimensions of testing programs.
- Resourceful and able to resolve project challenges.

We will maintain a schedule that ensures the customer service line is supported by a “live” person between the hours of 8 a.m. and 5 p.m. Eastern Standard Time. DRC acknowledges that this core time will be extended as necessary during peak customer service periods (e.g., one week before the testing window begins through one week after the close of the testing window). Our customer service function is organized such that only **staff trained in the GCA Program will respond to calls**. As part of our training process, a program-specific customer service manual will be in place and will include frequently asked

questions and responses, a program overview, and information on due dates, etc. This manual will play a pivotal role in standardizing the customer communication process for this program. The use of this manual will ensure that Pennsylvania School and District Assessment Coordinators who call DRC receive accurate and consistent information from DRC Customer Service staff.

DRC recognizes and understands the importance of being able to receive and respond to each call in a timely manner. DRC has extensive experience managing caller traffic, working with such states as Alaska, Idaho, Louisiana, Pennsylvania, and South Carolina. Unlike many testing companies that employ automated calling systems, which can strain the patience of school personnel, DRC assures PDE that a trained member of the GCA Project Management Team will answer all calls directly. **We have enjoyed getting to know Pennsylvania school and LEA personnel and Assessment Coordinators and welcome the opportunity to continue to support PDE's assessment initiatives, including the GCA Program.**

Toll-Free Customer Service Number and Other Communication Links

To respond to Assessment Coordinators' questions or concerns, DRC will provide reliable and timely customer service support. The communication links will be available throughout the duration of the contract and will include **a GCA-dedicated toll-free customer service telephone number, email address, and fax number**. These communication links will provide, at a minimum:

- A method for answering Assessment Coordinators' questions about enrollment and registration, materials delivery and collection, inventory methods, test administration, and packaging materials for return.
- A process for LEAs to order additional materials as needed after the original shipment.
- A system for tracking the delivery and return of materials and the delivery of reports.

Our team will provide consistent and immediate response to callers through the use of a customer service database. This database will track all customer contacts, from the initial call or email through the resolution of the issue or question.

"Data Recognition Corporation has provided a wonderful service to the educators of this state. It has been a 'user-friendly' organization that understands the needs of its clients. In serving as the Wilkes-Barre Area School District's assessment coordinator, I have found my association with DRC to be professional, friendly, and productive. I have gotten to know each of DRC's customer representatives on a first name basis and have never had a bad experience with any of them."

—Deputy Superintendent,
Wilkes-Barre Area School District,
Pennsylvania

DRC's highly experienced and trained Pennsylvania Customer Service Team members will be available to answer calls on the toll-free number from 8 a.m. to 5 p.m. Eastern Time each business day. All standard issues will be resolved within 24 hours; callers with complex situations requiring additional time for resolution will receive regular updates on the status of their issues until resolution is complete.

DRC has extensive experience managing high caller traffic, working with such states as Alaska, Idaho, Louisiana, South Carolina, and Washington. During peak assessment periods, DRC often receives over 2,600 calls each week. Each of these callers receives **prompt, responsive, personalized service** from our Customer Service staff. DRC is accustomed to providing extended hours customer service support during test administration windows. To meet the requirements of our Alaska client, during certain times of the year, we have dedicated Customer Service Representatives available to take calls until 7 p.m. Central Time. **DRC's dedication to ensuring customer satisfaction makes us willing to assist callers outside of normal business hours.**

In the unlikely event that telephone service is interrupted, DRC will send an email notifying assessment coordinators that the telephones are down and will send another email once service has been restored. In addition, DRC Customer Service Representatives will have access to cell phones that can be used in emergency situations.

Leaving Messages for Customer Service

As part of our dedication to customer satisfaction, DRC believes it is essential to respond quickly to all calls. In extremely rare cases of particularly high volume, if no representatives are available, callers will be able to leave a voicemail message. School and District Assessment Coordinators and PDE staff will also have the option of contacting DRC's Pennsylvania Customer Service staff through email and fax. DRC will make every effort to respond to messages within one hour. DRC will train additional personnel to respond to customer service inquiries during and in the weeks surrounding the testing window. This will ensure that we are able to respond to all inquiries in a timely manner.

Customer Service Email Support

DRC believes that email can be a useful form of communication in some circumstances. A GCA-dedicated email address will continue to be available for School and District Assessment Coordinators to use as another means of customer service support. Emails will be responded to in a timely manner. DRC Project Management will also use email to mass disseminate information to district assessment coordinators as needed. Mass communication with districts will be submitted to PDE for approval prior to being sent.

Customer Service Database

To enhance the services DRC provides to our state clients and state assessment coordinators, we have developed a centralized project information repository and database, known as the Education Project Information Center (EPIC). This proprietary system serves several purposes, including customer service call logging, call log reporting, district/school contact information storage, and shipping information maintenance. Below, we present an overview of the EPIC system and have provided several sample screenshots/reports.

EPIC will be integral to our customer service function. It will be used to track all customer contacts, from the initial call, fax, or email up through issue resolution, providing consistent and immediate response to callers. EPIC is password-protected; access to information is based on user roles defined specifically for each project. Highly configurable, EPIC is easily tailored to meet all of PDE's needs.

DRC's Customer Service staff will enter all communication into EPIC. Date and time stamps are automatically added to each EPIC entry. The information contained in the database is searchable by a wide variety of criteria—which can be customized for each project—such as district name, address, date, caller name, and resolution status (unresolved vs. resolved), as well as issue type, such as missing materials, material shortages, and precode processes. Search results can be sorted and narrowed using sort filters. If appropriate, each call can also be associated with more than one project for a state. EPIC provides the ability to track and monitor the resolution of every call; calls requiring follow-up and

resolution are flagged, ensuring that every caller will receive timely issue resolution.

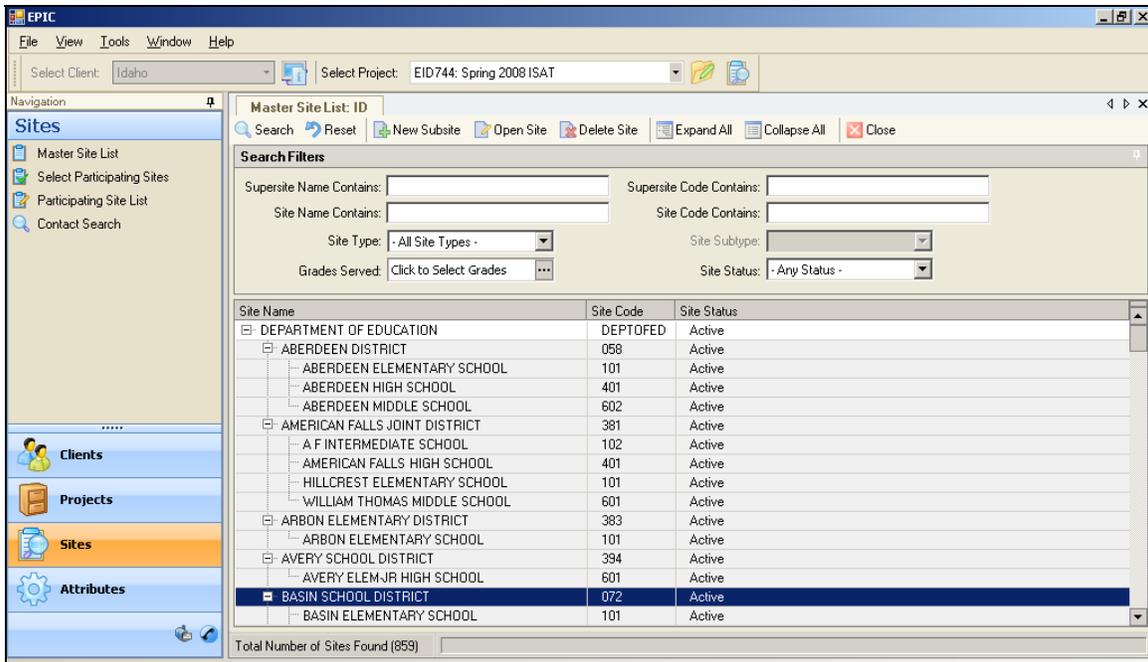
Shipping accuracy is also enhanced through the use of EPIC. Shipping addresses and local assessment coordinator contact information for each participating school and district are stored and maintained in EPIC. EPIC system-checks validate all entries created to ensure accuracy. EPIC also verifies that shipping addresses are valid prior to shipping. Then, at the time of packaging and shipping, DRC's proprietary Ops MMS system interacts with EPIC to utilize accurate, current shipping information. This allows for the incorporation of real-time shipping information updates for the most accurate site information and efficient process.

EPIC will provide the following benefits to PDE and Pennsylvania Assessment Coordinators:

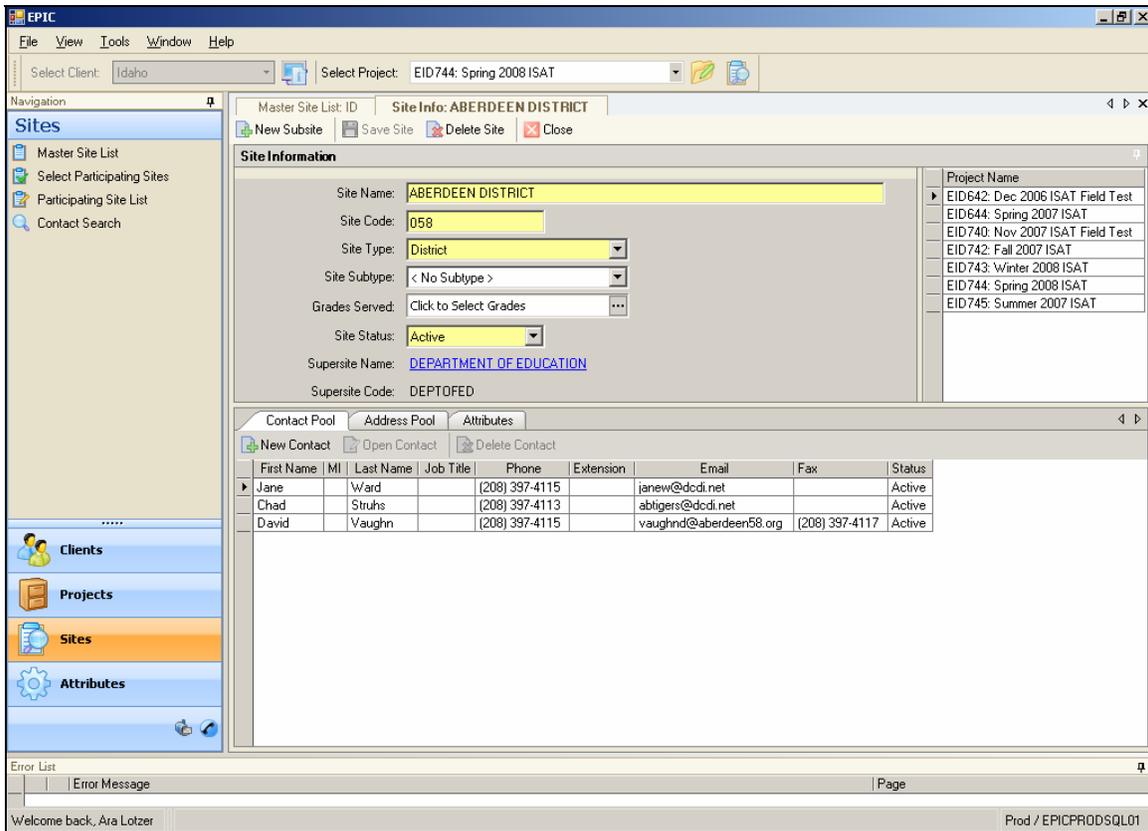
Benefits of DRC's EPIC Customer Service Database

- **Enhanced Service**—One central customer service database for each testing program ensures that every customer service team member has access to each logged call, providing consistency of service throughout the program's duration. The ability to flag, track, and log follow-ups ensures that all callers receive timely resolution of their issues.
- **Call Reporting**—Data fields are configurable for each program, allowing us to provide customized call reports to meet the unique needs of each of our clients. Because all calls are logged and categorized in EPIC, call reporting is real-time and accurate.
- **Data File Uploading**—Shipping and contact data provided by departments of education and other contractors/subcontractors are easily uploaded into EPIC. This feature helps maintain consistent and accurate shipping and contact information even during program transitions. Built-in checks validate all uploaded information.
- **Shipping Accuracy**—All contact and shipping information is stored and maintained in EPIC. Our Ops MMS system uses the information stored in EPIC to ensure accurate, real-time shipping information, eliminating the potential for error associated with maintaining multiple address databases. Contact and shipping data updates are automatically updated across all programs for a particular state as appropriate.

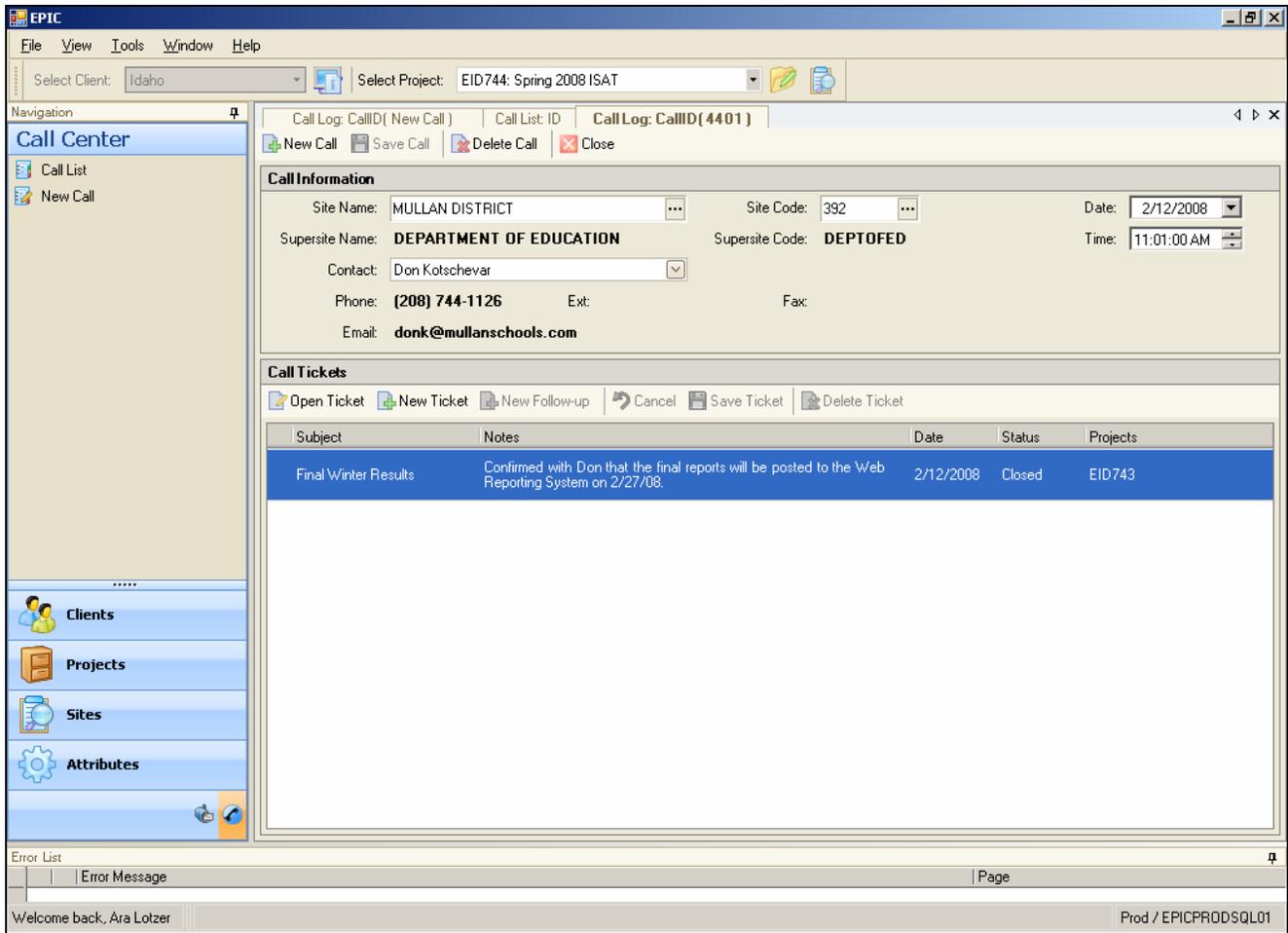
Sample screenshots are provided on the following pages, illustrating EPIC's master site list, contact information, and call logging functions.



EPIC Sample Screenshot: Master Site List of Districts/Schools



EPIC Sample Screenshot: Contact Information for a Site



EPIC Sample Screenshot: Call Log Screen

Call Log Reporting

EPIC’s configurable data fields allow us to provide customized call reports tailored for each of our clients. Because all calls are logged and categorized in EPIC, call reporting is real-time and accurate. Reports can be generated based on issues critical to each program, using issue categories selected by a particular client.

Customer Service Testimonials

DRC has earned a reputation for providing a level of customer service that we believe is superior to any other testing organization. **Our clients appreciate our dedication, commitment, and quick responses;** Alaska, Idaho, Louisiana, Pennsylvania, and South Carolina assessment coordinators will readily attest to our client-focused, professional, and personal service. The following quotes are representative of the numerous comments DRC has received from district staff.

Pennsylvania

“DRC members deserve outstanding praise for their remarkable knowledge of the assessments, their amiable and patient support of their clients, their professional decorum, and their conscientious commitment to their tasks. Many of my colleagues also affirm that DRC’s support with the PA Assessment process makes it efficient and effective.”

“At all times, DRC employees are not only professional and knowledgeable, but to list only a few qualities, they are polite, patient, hard-working, good natured, judicious, and ethical.”

“Since my first contact with DRC years ago, I can say nothing but good things about the reception that I have always received from your staff. From my questions, concerns, etc., they have always been kind, courteous, and treated me as if my question was the most important one. I strongly believe that each of your staff is extremely well educated in the rules/guidelines of the GCAs testing, as it doesn’t matter who answers the call, they have the answer.”

“We were very appreciative of the promptness of DRC returning any and all phone calls we made. DRC was very helpful in getting our questions answered. They were thorough and very pleasant to speak with.”

“Your support team should be commended. They are always polite, prompt, and precise with answers to any questions.”

Alaska

“Again—thanks for making my life so much easier. The packaging of the materials is fantastic. The other nice thing is that all of the additional ‘stuff’ is in the first box and not in the box where it may have fit the best for those doing the packaging. In the past, we have searched through box after box to find the school inventory information. Nice job!!!!”

“I love the service I get from DRC. Thank you so very much.”

“Thank you so much. It is great to be able to trust that this was going to get done. You have a great reputation amongst the [district test coordinators] up here. It’s pretty cool.”

“For the record, I just wanted to let you all know, I have been nothing less than very impressed and appreciative with the services DRC has provided these last three years. Thank you!”

“Thank you for your help which is above and beyond the call of duty. I appreciate your professionalism in making this a very easy transition of information. If the rest of my job would go as well as this has I could say it would be a breeze.”

Idaho

"I wanted to take a moment to thank all of you for a job very well done. Your team has been incredibly helpful and worked very well with us to make everything work efficiently. I know you do not hear this nearly enough, but you folks are doing a great job and should be commended."

"I really do appreciate these responses—I can communicate more effectively with my administrators and teachers to help them more fully understand student scores."

"Thank you for answering my question so efficiently and effectively."

"Thank you so much for your speedy replies. They are truly appreciated!"

"Thank you SO MUCH for your help in resolving our issues...you served your company well yesterday."

Louisiana

"You are so pleasant and a joy to work with. It makes my job so much easier to have you there to assist me!"

"Thanks for the constant/consistent support and availability."

"It was a pleasure working with you the last couple of years. You were really helpful in resolving problems concerning assessment materials."

"Thanks for prompt answers to email and returned phone calls."

"No one at DRC makes me feel stupid, no matter what questions I'm asking about the assessments!"

South Carolina

"I want to thank you and your staff for the wonderful service you provided this year. Your prompt and courteous responses to my calls and emails is much appreciated!"

"Members of the South Carolina Project team displayed great patience and provided such outstanding customer service, and because of their efforts, my program ran smoothly from start to finish without a major glitch."

"DRC seems to have streamlined the process—the staff is wonderful to work with."

"I want to let you know just how grateful I am for you and for all of the support that you have given me this year, especially since I am a new coordinator. Your patience, understanding, and quick response have helped me to adjust to my new position. Knowing that I can call on you and the service that you give me make my job easier. Thank you."

"You all are such professionals ... You all truly care about your customers and I can't say that about many companies in the corporate world. Thanks again, you made my day!!!"

VII.G.2. Assessment Administration Procedural Workshops

In cooperation with PDE, DRC will develop, plan, and conduct web-based test administration procedural workshops throughout Pennsylvania in accordance with PDE's requirements and approved schedule. We will provide these workshops for LEA and school/testing site test coordinators on an annual basis in preparation for GCA administrations.

Additionally, DRC will take direction from PDE regarding critical issues and effective methods for the web-based training sessions. We have extensive experience providing this training to the LEAs and site test coordinators in Pennsylvania for the past five years for the PSSA and DRC provides similar training for many of our other large-scale assessment clients. We look forward to working with PDE to develop the training and preparation processes for the GCA Program.

PDE-approved announcements and registration information for the training will be broadcast to LEA/school testing coordinators. The registration information will also be posted on the PDE website.

Each training opportunity will address the procedures, materials, and timelines associated with GCA administration activities. Specific training topics will be selected in consultation with PDE. Once the topics have been identified, DRC will organize and execute all steps necessary to conduct the training. DRC will develop and submit all training materials, including any PowerPoint and video presentations, to PDE for review and approval prior to each session. If necessary, we will also prepare reference materials (e.g., sample demographic pages) and produce electronic versions for electronic distribution and/or posting online. Please see *Appendix 5* for a sample training handout.

DRC trainers will be DRC Pennsylvania team members experienced with GCA assessment administration. We will ensure that every web-based workshop presents the same information. Workshop participants will leave with the confidence that they can readily obtain additional support from DRC's Pennsylvania Customer Service Team. Additionally, follow-up materials (such as brochures and other handouts) will be made available electronically, as approved by PDE. The availability of presentation and training materials online will be useful for new staff hired by LEAs/schools after the training sessions.

DRC's experienced IT personnel will work alongside the DRC GCA Project Management Team to implement security precautions and provide technical support. PowerPoint and video training materials can all be integrated into web-based training workshops. During these training sessions, DRC staff will be available online to answer questions from participants. Participants will be able to interact with the training through an online chat feature.

DRC's Project Management Team will direct, monitor, and ensure that the following activities are carried out for the web-based training sessions:

- Obtain PDE approval for all training materials.
- Schedule web-based workshops, approved by PDE.
- Post training schedules on the PDE website.
- Distribute announcements and registration forms to LEA/school testing coordinators and post the registration information on the PDE website.
- Prepare and produce workshop materials for participants, in collaboration with PDE.
- Electronically distribute materials to workshop participants.
- Coordinate with PDE prior to and during workshops.

DRC is pleased to conduct web-based assessment administration workshops for the GCA Program. However, we would be very interested in discussing the possibility of providing in-person (face-to-face) training workshops, along with PDE, at various locations around Pennsylvania upon contract award. If PDE is interested in this option, DRC would provide cost information.

VII.G.3. Procedures for Maintaining Assessment Material Security

DRC understands that ensuring security is critical to maintaining the technical quality, perceived fairness, and integrity of any testing program. We recognize that assessment security is of the utmost importance to PDE. We have proven quality-control and security procedures integrated throughout all of our operational processes. We will also implement other necessary security measures as requested and approved by PDE to enhance security.

Please refer to *Subheading IV.C.15., Security of Test Materials and Results*, for more information on DRC's security features and procedures.

VII.H. ASSESSMENT MATERIALS COLLECTIONS AND SHIPPING

VII.H.1. Procedures for Collection and Receipt of Assessment Materials

Collection of Student Test Booklets

In order to accommodate the turnaround times requested by PDE, DRC is proposing that all test materials are returned to DRC directly from schools using UPS regardless of ship-to-school status. DRC feels that this will simplify the return process immensely for school and district personnel.

The School Assessment Coordinator has the final responsibility for the assembly of testing materials, the return of all secure materials from individual administrators, and preparing school materials for collection. DRC's return system fully enables School Assessment Coordinators to account for all secure documents. Administration Manuals and Assessment Coordinator's Handbooks will contain explicit instructions for maintaining security and handling and packaging materials for collection. School Assessment Coordinators will be encouraged to **call DRC's toll-free customer service number** should questions occur during packaging.

After each assessment, test administrators will package secure materials and return them to the School Assessment Coordinator. The School Assessment Coordinator will verify the return of all secure test materials from all classrooms using the School Security Checklist, which School Assessment Coordinators will keep for their records.

After verifying that all secure materials were returned from test administrators, School Assessment Coordinators will package all school boxes for return to DRC. All materials will be returned to DRC via UPS.

Materials Return Shipment

To enhance the accuracy and decrease the turnaround time of the return shipment process, DRC provides a clear, well-documented return process in the Administration Manuals, Assessment Coordinator Handbooks, and Return Shipment Instructions. We offer School and LEA Assessment Coordinators support from our Pennsylvania Customer Service staff, which has a history of providing superior service to Pennsylvania LEAs, schools, and other testing sites for the PSSA (please see *Subheading VII.G.1., Customer Service Support*, for more information about this function).

Unlike many other testing companies, our receipt processes require little document preparation by schools and districts. Additionally, we do not require special packaging or return procedures for accommodated materials. Our Ops MMS system enables 100% accuracy in accounting for returned barcoded materials regardless of how materials are packaged or bundled.

DRC will use UPS for the return of all test materials. UPS Return Service (UPS RS) will be used for scorable materials; UPS (ground service) will be used for the return of all nonscorable test materials.

Upon receipt at DRC, all returned boxes will be scanned in through our automated Box Receipt System. Received materials are compared against the shipper's manifest and testing site counts to identify any discrepancies, which are forwarded immediately to Project Management for resolution.

Materials Return Quality Procedures

- **Clear instructions for assessment administrators and coordinators—** Assessment Coordinator Handbooks and Return Shipment Instructions will provide straightforward, easy-to-follow instructions for handling and packaging materials for collection and return. DRC's exceptional customer service support will be available between predefined business hours.
- **Control Forms—** School Assessment Coordinators will verify the return of all secure test materials from all classrooms using the School Security Checklist.
- **Traceability—** The shipping carrier (UPS) used for materials return will have an online, traceable distribution system to track all materials to provide the status of each shipment from the time it is collected and returned to DRC.
- **Tracking of boxes—** Upon receipt of materials at DRC, all returned boxes will be scanned in through our automated Box Receipt System. Quality-control reports are generated to compare materials received against the shipper's manifest and the district counts. Materials return information will be reported to PDE on a daily basis.
- **Tracking of test materials—** After box receipt, test materials will be separated for processing using DRC's Operations Materials Management System (Ops MMS). Any discrepancies in expected counts of materials based on original packing will be reported to Project Management for resolution.
- **Missing Materials Reports—** DRC will generate Missing Materials Reports, which will be available for PDE to review. After all materials have been checked in and discrepancies have routed for resolution, a final report will be generated for PDE.
- **Communication with PDE—** DRC's Project Management staff will communicate with PDE regularly during the entire materials receipt process to discuss any concerns or issues.

Please see *Subheading IV.C.15., Security of Test Materials and Results*, for DRC's materials receipt and secure processing procedures.

VII.H.2. Providing Supplies for Return of Materials

DRC will provide all school-specific return shipping labels, forms, and boxes. We will be responsible for all costs associated with the return of materials. DRC encourages schools to re-use boxes from their original shipment for the return shipment, but we will send additional boxes upon request.

VII.H.3. Communication with District Coordinators

Schools will call UPS to schedule pickup of scorable test materials (i.e., answer documents or scannable test booklets). DRC will present clear instructions for contacting UPS and will provide toll-free numbers and/or website addresses. UPS uses an online tracking system that provides the status of each shipment from the time it is collected until it is returned to DRC.

VII.H.4. Postage and Shipping Costs

DRC will be responsible for all distribution and retrieval costs. Please see *Subheadings VII.F., Materials Delivery*, and *VII.H., Test Materials Collections and Shipping*, for more a discussion of our materials delivery and collection procedures.

VII.H.5. Check-in Procedures for Receipt of Materials

DRC achieves accurate and efficient secure materials receipt processes that require a **minimal amount of document preparation on the part of schools and districts**. Our proprietary materials management system, Ops MMS, allows us to accomplish this goal. Its advanced automation and barcode scanners provide fast and accurate data collection with **no dependence on materials/document order**. This translates into **time saved for assessment coordinators** during materials return. Captured data are organized into user-friendly reports from the start of the secure materials check-in process, providing valuable insight into suspected material shortfalls as early as possible, and mitigating potential consequences of delay.



As evidenced by our **ISO 9001:2000 certification**, DRC maintains stringent quality-control procedures during the document receipt process. Log-in procedures, developed by the GCA Program Manager and Document Processing Manager and approved by PDE, will provide our clerical personnel with step-by-step instructions to be followed during the log-in process for GCA secure materials. DRC uses Ops MMS to provide efficient and accurate control of all secure documents that are shipped from and returned to DRC for each test administration. Ops MMS will identify all secure test materials by site code and provide an automated quality check between items designated for and returned from a site based on the following, at a minimum: name of testing program, site (school/LEA), course, material type, and quantity. Using barcode technology, the system allows us to track the documents through all operational steps: box receipt, materials separation, booklet check-in, scanning, editing, data validation, and data conversion for reporting.

All returned boxes will be scanned in through our automated Box Receipt System. When the box receipt process is complete, pallets will be retrieved (according to the specified priority for processing) from their staged positions in the warehouse racking and delivered for the materials separation phase. Warehouse personnel will open the boxes from these pallets and sort the contents by course, status (i.e., used/unused), and any other GCA-specific criteria.

Project Management will provide warehouse personnel with complete instructions/training (previously approved by PDE) for manual sorting and prioritization of returns. Project Management will monitor the sorting and check-in of test materials daily. Warehouse personnel will alert Project Management of

any test booklet and/or answer document that is not returned in the specified manner for each return shipment. The GCA Project Management team will communicate with PDE staff regularly during the entire materials receipt process to discuss any concerns or issues related to secure material return and processing.

When materials for a school are completely sorted, each type of sorted material will be boxed separately. Once filled, a sorted box will be staged for document counting. Next, the documents are loaded into an automated counter which, working in conjunction with the Ops MMS, will piece count each document in the box. This count will remain correlated to the box in the system, providing a target number and an essential quality-control step for the booklet check-in process. An on-demand label will be produced that contains a barcode representation, as well as a human-readable description of the material contents and quantity. Other data will also be captured by Ops MMS at this time to facilitate numerous internal quality-control functions. Once labeled, the sorted and counted box will be staged for the next step.

During the next phase, sorted material will be retrieved and sent through DRC's booklet check-in system, which uses streamfeeder automation, coupled with Ops MMS, to collect and carry documents past oscillating scanners that capture data from up to two representative barcodes. For the GCAs, we are proposing that secure test booklets follow the booklet check-in process, but answer documents do not. The following describes this process:

- The check-in operator uses a hand scanner to scan the label of the sorted, counted box and loads the contents into the Streamfeeder. This initial scanning of the label inputs the parameters of what material type and quantity Ops MMS should expect in the forthcoming box of documents.
- The documents in the box are then fed past the scanners, which capture the security number and the precode number simultaneously.
- Once scanning begins, the operator monitors an Ops MMS screen. A display of scan errors, as well as an ordered accounting of what has been successfully scanned, will appear, along with the document count for that box. Using this mechanism, errors can be quickly detected and resolved.
- When all materials in a box have been scanned and the pre-established document count has been reached, the box is sealed and placed on a pallet by material type. If the count is not reached or if the operator encounters difficulties that he or she cannot resolve, the box and its contents are delivered to an exception handling station for resolution.
- A daily exception report is produced to account for boxes completed through booklet check-in that did not reach the document count established in the counting phase. Any boxes on this report will be recovered and resolved as part of exception handling.
- All captured data are stored in the Ops MMS database.

- Completed pallets are staged in warehouse racking until required for further processing.

Given that this check-in process must occur immediately upon receipt of GCA materials, DRC Operations staff will be able to provide real-time feedback to the GCA Project Management Team on actual receipts versus expected receipts from schools. In turn, the Project Management Team will contact any school regarding what appears to be an anticipated materials receipt “shortfall” as soon as the materials for the entire school are checked in. Secure materials issues can thus be identified and resolved well in advance of any reporting. The GCA Component Lead will communicate with PDE staff regularly during the entire materials receipt process to discuss any concerns or issues related to secure materials return and processing.

Our processing system, which is ISO 9001:2000 certified, offers a tremendous advantage to PDE by providing quality-control measures that are specifically related to potential test security issues. Problems can be caught early and resolved in a timely manner.

VII.H.6. Missing Materials Report

DRC recognizes that the security of the test is of the utmost importance to PDE. To that end DRC has implemented several processes that will help ensure the security of the test booklets. Because each test booklet has its own unique test security number on it, after booklet check-in for the entire project is complete, DRC will provide PDE with a report containing missing documents by security number, as well by course and school and other criteria required by PDE.

DRC will provide a Missing Materials Report to PDE within 20 days from the end of each testing window. After school staff have located missing materials or confirmed that secure materials are not in the school, DRC will produce the final Missing Materials Report. If desired by PDE, this report will be produced for each LEA and each school within the LEA, along with a consolidated statewide report. Any materials not returned to DRC can be listed by security number, type of document, course, etc., depending on PDE requirements. The security report will list the number of materials not returned and the LEA(s) to which they were originally sent, as well as summarize any problems noted during materials return/check-in, based on PDE requirements. These reports will assist DRC and PDE in improving the instructions in Administration Manuals and Assessment Coordinator’s Handbooks, as well as information shared in the online administration procedural training sessions. Reports will be produced from an error log maintained by Project Management while handling problems that may arise during materials return. In addition, DRC will provide PDE with complete documentation of the steps that were taken by DRC and the schools to locate any missing secure materials.

VII.I. SCANNING/IMAGING AND SCORING

VII.I.1. Equipment and Software

With 30 years of impeccable service, DRC's Operations Department provides:

- Client-tailored processes and solutions.
- Reliable and efficient systems.
- Adherence to stringent quality-control procedures.

We believe that our combined experience and expertise in the areas of materials processing, image scanning, and performance assessment scoring is unsurpassed in the industry. DRC's 30 years of experience, including 16 years of Pennsylvania experience, has resulted in a scannable document process that is extremely reliable and efficient; our processing and scanning processes are ISO 9001:2000 certified. Our Operations personnel take pride in our ability to tailor processes to meet each of our clients' needs. All

processing and scanning occurs at fully secure facilities. DRC has successfully processed millions of scannable test materials for large-scale statewide assessments, including the following:

- Alaska
- Louisiana
- Oklahoma
- Pennsylvania
- South Carolina

In the past seven years of image scanning and scoring, DRC has met our internal handoff deadlines and has successfully delivered results for our clients. All of our statewide assessment contracts utilize image scanning and scoring. In the spring of 2007, DRC image scanned and scored more than 4.5 million student answer documents consisting of more than 100 million pages (50 million sheets). In spring 2008, the processing quantities were even higher, with Pennsylvania PSSA science fully operational. For the Pennsylvania PSSA alone in 2008, we processed a total of 1,950,000 student answer booklets, with a total of nearly 60 million pages (30 million sheets).

The proposed Operations staff listed below has extensive direct experience with multiple statewide assessment programs.

Mr. Doyle Kirkeby, Vice President of Operations, will provide senior management support to complete all operations activities for the GCA Program, as he has for the past thirteen years, including scanning and editing—with precision and timeliness. Mr. Kirkeby has served as a senior leader in DRC Operations since 1995 and has supported Operations staff and functional areas in accessing additional training and technology to increase efficiency and achieve high-quality standard operating procedures.

Ms. Ginny Burnett, Senior Director, Operations—Education, will oversee all scanning and editing activities for the GCA Program. She has 20 years of experience in operations activities, including those for numerous large-scale

assessment programs. She has implemented quality assurance procedures in the Operations environment at DRC, expanded the use of technology as it applies to clerical quality functions, maintained updated clerical systems, and provided staff training on new hardware and software.

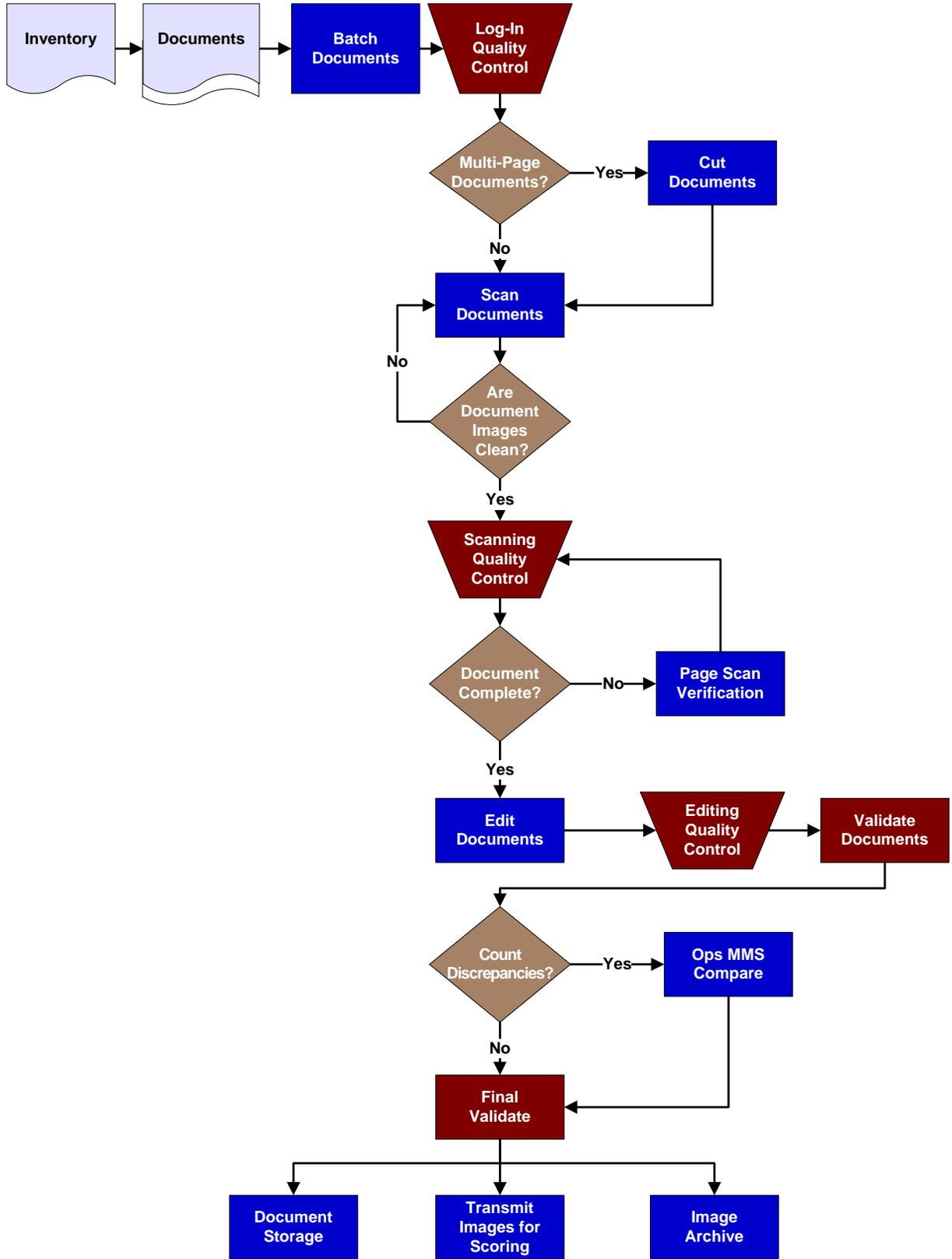
Document Scanning

DRC's state-of-the-art proprietary scanning system is highly configurable and fully scalable, which provides **the flexibility needed to accommodate the GCA Program and each of our other state client's needs**. DRC's customized scanning programs are capable of selectively reading documents and electronically formatting scanned information. The IBML ImageTrac scanners can process single- or multiple-color documents. All custom scanning programs go through quality review before testing materials arrive. Our image scanning operators have extensive experience performing tasks related to scanning and the maintenance of image scanning equipment.



During the scanning stage, open-ended (OE) items will also be captured as images and loaded into the Image Scoring System for handscoring. The **use of image scanning and scoring technology** at DRC **mitigates the inefficiencies of traditional paper-based packeting** of OE items and student responses.

A high-level overview of our scannable document processing is shown in the following figure.



Scannable Materials Processing

VII.I.2. Plan for Scoring of the GCAs

DRC has 30 years of experience scoring and reporting large-scale assessment results. Our extensive reporting experience for the PSSA and other assessment programs such as those for Alabama, Alaska, Louisiana, South Carolina, and Washington, can assure PDE that DRC has the ability to score and report accurate results in critically prescribed time limits. For a full discussion of our proposed reporting plan, including timelines to meet the requirements of the RFP, please see *Subheading VII.K.*

VII.I.3. Quality Assurance Plan

DRC is passionate about providing quality products and services to our clients and recognizes that quality processes are critical elements of our business. Quality at DRC is being taken to world-class levels, providing us with yet another competitive advantage.

A primary factor in DRC's continued success in providing error-free services to clients is our company-wide dedication to quality.

With nearly 30 years of successful student achievement testing, we have developed and refined our quality system to ensure the highest levels of customer satisfaction and quality. **At DRC, quality is both a program and an overall approach to business.** Our Quality Management System is focused on defining and implementing critical quality control processes to ensure products and services delivered to our clients meet and exceed their requirements. This extends to our relationships with other vendors and partners.

At DRC, quality is a commitment to excellence and is achieved by teamwork and the process of continuous improvement. Quality principles are infused into everyone's roles within our organization. We are dedicated to being the quality leader in the industry and are confident our products and services will exceed PDE's expectations. The focus of our Quality Management System is to define and implement quality control processes and embed them throughout all aspects of our projects. DRC has developed our quality approach using the guidelines listed in the *SCASS/TILSA Quality Control Checklist for Processing, Scoring, and Reporting.*



Quality begins with the attitude that each task must be done right the first time. DRC staff members take great pride in their work, and their products reflect that pride. As demonstrated on all current DRC contracts, we understand the tasks that are necessary for successful assessment programs. DRC believes in reasonable and sensible approaches to problem solving. We pride ourselves on our creativity and our ability to anticipate problems, as well as our genuine affinity for discovering multiple solutions to difficult issues. DRC is eager to maintain a mutually satisfying relationship with PDE.

ISO 9001 Certification

In today's world, customers continue to evolve their wants and needs. They are asking us to be more dynamic, flexible and cost efficient in meeting their requirements than ever before. This places a tremendous amount of importance on our processes to meet these needs in a reliable, repeatable fashion.

That is why DRC made the decision to attain ISO 9001 certification in 2007. ISO 9001:2000 is an internationally recognized quality management standard that defines a set of core quality requirements an organization must comply with. Some of the requirements in the ISO 9001:2000 standard include:

- A set of procedures that cover key processes within a business.
- Monitoring manufacturing and business processes to ensure they are producing quality products and services.
- Keeping proper records.
- Checking outgoing product for defects, with appropriate corrective action where necessary.

- Regularly reviewing individual processes and the quality system itself for effectiveness.
- Facilitate continual improvement customers expect.

DRC is currently ISO 9001:2000 certified in three major areas of the company:

- Document Services (Project Management, Publications, Pre-Press, Printing, Bindery, Inserting, and Purchasing).
- Educational Operations (Distribution, Logistics, Materials Processing, Warehousing and Document Scanning).
- Woodbury and Minnetonka, Minnesota and Cincinnati, Ohio Scoring Centers.

External validation from a third party is required for a company to become ISO 9001 certified. An organization known as a “registrar” evaluates whether DRC is meeting the criteria of the ISO 9001:2000 standard within our quality management system. These “audits” are conducted twice annually.

The scope of our ISO 9001:2000 registration is based on a “business process”, rather than a “functional” approach that many companies apply. Embedding the ISO 9001 standard has enhanced an already strong foundation of business process controls that has been DRC’s hallmark for many years.

Our ISO 9001 certification process is led by **Mr. Niall Finn, Director of Quality for DRC’s Operations**. Mr. Finn has extensive hands-on quality management experience in various manufacturing environments. As the senior quality leader responsible for leading the implementation of ISO 9001 Quality Management System certification across all DRC operational areas, he will continue to oversee the plan to expand the scope of our certification to other areas of the company, while contributing his expertise to our quality standards and systems already in place.

Project Management and Planning

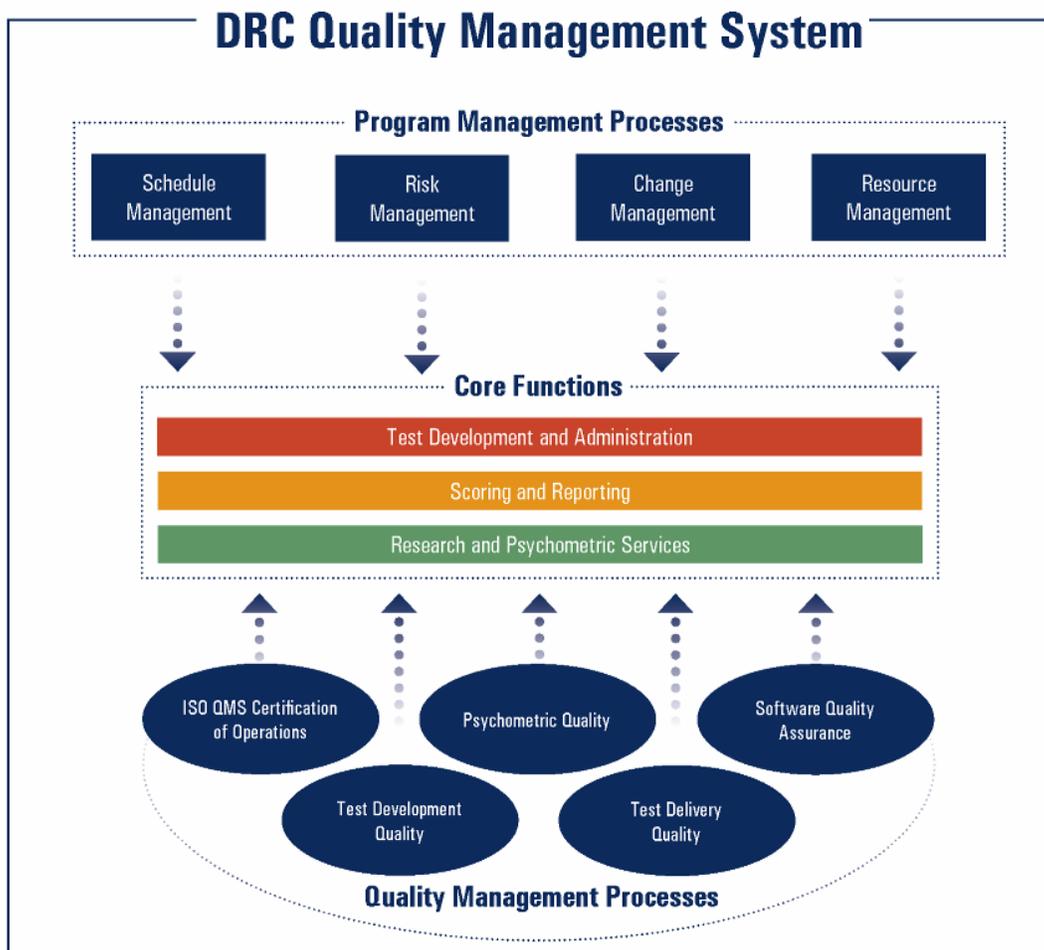
For the success of the GCA Program, PDE’s requirements, goals, and constraints must be thoroughly understood, documented, and communicated. These critical activities are the foundation of DRC project management activities. **Dr. Adisack Nhouyvanisvong, GCA Program Manager**, will be responsible for the administration of the overall quality process. Problem-reporting procedures will be strictly followed to ensure immediate action is taken to resolve any issues.

DRC’s **Vice President of Quality, Ms. Lisa Peterson-Nelson**, will also carefully audit the project delivery process for the GCA Program. She is currently directing the enhancement of DRC’s key work processes for delivery of products and services to clients. Ms. Peterson-Nelson has over 19 years of experience in quality process improvement. She worked for more than a decade in senior positions in

quality process management for two different Fortune 500 companies. She has been with DRC since 2001.

We will provide PDE with the required evidence that our quality inspections, processes, system tests, and policies are followed. In addition, DRC will also provide PDE with a Quality Control Manual at the end of each contract year detailing the quality procedures used throughout all phases of the project. The manual will be updated yearly and will include any changes in processes or procedures.

To ensure the success of the GCA Program, we will proactively manage risks, such as programmatic, technical, cost and schedule risks. Dr. Nhoyvanisvong will function as the risk manager by working with other members of the project team to provide the work breakdown for the project, develop detailed scope of work agreements, and design of a formal risk management matrix specific to the project. Dr. Nhoyvanisvong will schedule and oversee risk reviews, in conjunction with the GCA Project Team and PDE. Ms. Lisa Peterson-Nelson, DRC’s Vice President of Quality, will also provide support to the risk management process, providing an additional level of program security.

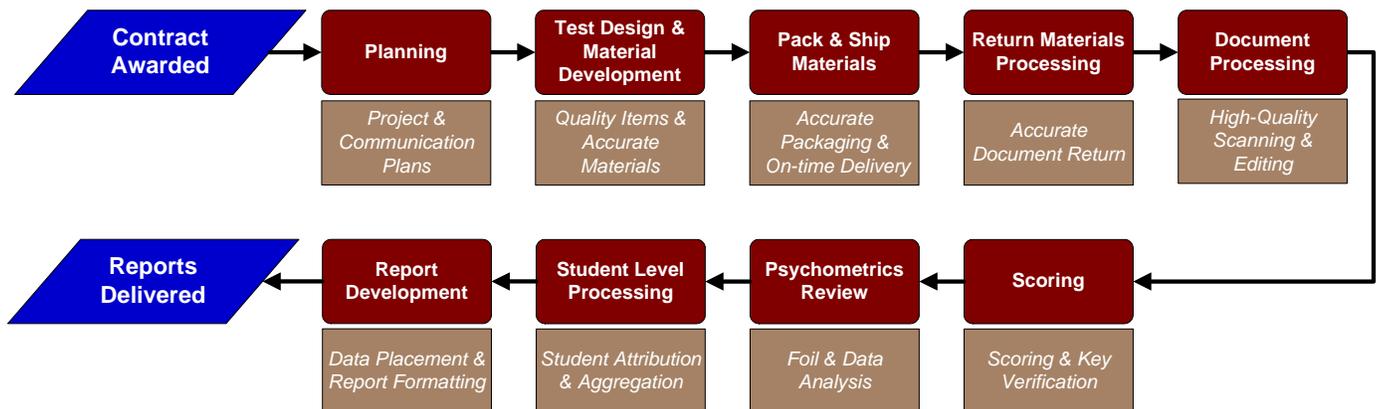


The GCA Program necessitates a partner that is flexible, innovative, and prepared to manage change. At DRC, change management is a critical management discipline. Our change management process is used to control and manage size, effort, cost, and schedules. Because change can occur at any time, we have implemented activities in our process to identify change, control change, and ensure change is properly implemented and reported to groups who are affected. PDE can be assured that DRC will thoroughly evaluate each requested change and perform a detailed impact and risk analysis. We will provide PDE with our recommended implementation plan and clearly outline any schedule or cost impacts.

Quality Control Process Overview

Our Project Delivery Quality Control process begins with the contract award and ends with the distribution of all required deliverables. Quality control checkpoints are in place at all stages of processing. Our proven quality framework is an integral part of ensuring accurate and timely delivery for our clients. We will provide PDE with the required evidence that our quality inspections, processes, system tests and policies are followed.

Our Project Delivery Quality Control process is illustrated in the figure below.



Project Delivery Quality Control Process

Test Materials Quality Procedures

DRC understands that state departments of education require error-free materials; We take great pride in the excellence of the state testing publications we produce on behalf of our department of education clients. GCA Program testing-related materials will be **produced through DRC's ISO 9001-certified Document Services Division** that incorporates our Document/Graphics Design Group and our complete in-house Printing Department. This incorporation of resources gives DRC a unique capability to customize our processes to address the requirements of each of our clients within restricted parameters and rigorous timeframes.

In addition to ISO 9001 certification, DRC has earned **Quality Level II status from the Government Printing Office (GPO)**, which is the second highest status that can be awarded (Quality Level I status is reserved for printers who produce bound books, four-color varnished promotional pieces, or other similar materials).

DRC collaborates with each of our clients to maintain a program style guide to ensure consistent application of preferences and expectations across all program materials. Documents printed at DRC are printed to exacting specifications to guarantee the highest possible data integrity for OMR, OCR, and Imaging machines.

DRC follows a meticulous set of internal quality standards to ensure high-quality printed products for its clients. DRC assures PDE of our commitment to produce accurate test materials.

Test Materials Quality Procedures

- **Publishing and Editing Review**—DRC staff will perform a three-way review of all project materials. This process includes multiple group checks of answer keys to verify accuracy. After this internal review, assessment materials will be forwarded to PDE for review and approval.
- **Taking the test**—Staff will take the actual tests to ensure that all items and passages perform as planned.
- **In-house printing**—DRC's in-house printing department will print scannable materials based on predetermined specifications for quality and accuracy. External printing companies hired to print nonscannable forms will need to guarantee DRC the highest level of quality. DRC Project Management will review the vendor's quality plan.
- **Multiple checks**—DRC Project Managers and Print Procurement Specialists will routinely conduct meticulous quality checks during the printing process to see that all requirements for printed materials are met.

Packaging, Shipping, and Materials Return Quality Procedures

Accurate packing, shipping, and collection of test materials is critical for districts and schools to successfully administer the GCAs. DRC is proud of our quality excellence in this area and are committed to upholding that level of excellence for PDE.

Packaging and Shipping Quality Procedures

- **Detailed instructions**—Based on contract requirements and specifications, detailed Scope of Work Agreements (SOWAs) will be established by the DRC Project Managers working in conjunction with our Operations staff. The SOWAs will be available for PDE review at each step of the process.
- **Walkthroughs**—The Project Management team will conduct a walkthrough of the assembly process prior to each shipment to check that all procedures are precisely followed.
- **On-going monitoring**—The Director of Materials Operations and Logistics Manager will monitor the materials assembly area and report any irregularities to Project Management.
- **Secondary checks**—Our Operations staff will perform secondary checks on all packing lists and boxes are sealed for shipping.
- **Easy identification**—All district and school shipping labels will be quality checked to prevent materials going to the incorrect location. Site labels on each box will be compared to the shipping address label and matched for accuracy.
- **Traceability**—Shipping carriers used have online, traceable distribution systems to track all materials.

Materials Return Quality Procedures

- **Tracking of boxes**—Upon receipt of materials at DRC, all returned boxes will be scanned in through our automated Box Receipt System. Quality control reports are generated to compare materials received against the shipper's manifest and the district counts. Materials return information will be reported to PDE on a daily basis.
- **Tracking of test materials**—After box receipt, test materials will be separated for processing using DRC's Operations Materials Management System (Ops MMS). Any discrepancies in expected counts of materials based on original packing will be reported to Project Management for resolution.
- **Missing materials reports**—DRC will generate missing materials reports, which will be available for PDE to review. After all materials have been checked in and discrepancies have routed for resolution, a final report will be generated for PDE.
- **Communication**—DRC's Project Management staff will communicate with PDE regularly during the entire materials receipt process to discuss any concerns or issues.

Scanning Quality Procedures

DRC's image scanning and hand-scoring system was designed and built to work for all DRC imaging projects. Having a common scanning and hand-scoring system and platform eliminates the need for significant software development efforts to scan and score new projects. If enhancements are required for a project, the Imaging Information Systems department follows the proprietary DRC

software development methodology to complete development. This methodology outlines the standard deliverables for each phase of the development lifecycle (analyze, build, test, implement). Prior to implementation, all enhancements are reviewed and verified by the Software Quality Assurance department (SQA).

Software Development Quality Assurance and Methodology

DRC is committed to embedding quality throughout every aspect of our software design, development, and quality assurance processes, ensuring 100% accuracy in our scoring and reporting systems.

DRC's quality values start with:

- Building quality into the software development process.
- Following a methodology that focuses on the prevention of software issues.
- Providing highly qualified and trained staff.

DRC's Senior Director of Information Systems Software Quality Assurance

will oversee all software quality assurance activities for the GCA Program. The DRC Software Quality Assurance Team, which is comprised of dedicated software quality professionals, is specifically trained in the following areas:

- Software Quality Assurance.
- Software Testing.
- Software Development Processes.
- Quality Process Improvements.
- Standards and Requirements.

Software Quality Assurance staff will apply industry-standard software quality assurance methodologies throughout the program. DRC quality plans will be developed and will be available for PDE's review, if desired. Software Quality Assurance staff follow our project delivery quality control process and adhere to the quality control checkpoints for processing, scanning, and editing, described by the State Collaborative on Assessment (SCASS) on Technical Issues in Large-Scale Assessments (TILSA). Our proven software quality assurance standards and procedures are directly aligned with the Capability Maturity Model (CMM) from the Software Engineering Institute (SEI).

Test Decks

Prior to any GCA Program test materials returning to DRC, the Software Quality Assurance staff will perform extensive tests to ensure all scanned data (including demographic and multiple-choice responses) are captured and

accurately stored in a secure database environment. Each record in the database will be independently verified against the test decks for validation.

The analysts will follow a software testing methodology that thoroughly evaluates and verifies the scanning and scoring system and verifies each scanner is configured and setup for the GCA Program. This process includes validating test decks, which will be comprised of answer documents with and without student and school pre-id information for each form/version of the test. The test decks will be specifically gridded to include a variety of possible student response permutations and combinations.

The test decks will be processed completely through DRC's systems to verify the following:

- Readability of security, student and school barcodes.
- Data capture of pre-gridded and barcode information.
- Accurate capture of district and school codes.
- Consistent data capture on all scanners.
- Accurate scan positions on all documents and forms.
- Scanner calibration and hardware functionality.

The Software Quality Assurance staff will also perform a validation of all production data processed through the system. **Each student record will be verified for accuracy** to ensure high quality data file development and reporting.

Scanning Quality Procedures

- **Test Decks**—DRC will process test decks configured for the GCA Program through the production systems.
- **Calibration**—Daily calibration and scanner cleaning processes will be conducted to ensure read level consistency.
- **Standard edit processes**—Every scannable document will be processed through edit programs to detect potential errors (double marks, smudge marks, omits, etc.)
- **Multiple reviews**—The Document Processing Supervisor will conduct a review of the entire first batch prior to full production to ensure error-free processing.
- **Quality control reports**—Daily quality control reports for each editor will be reviewed by the Document Processing Supervisor to monitor the accuracy of the online editing process.

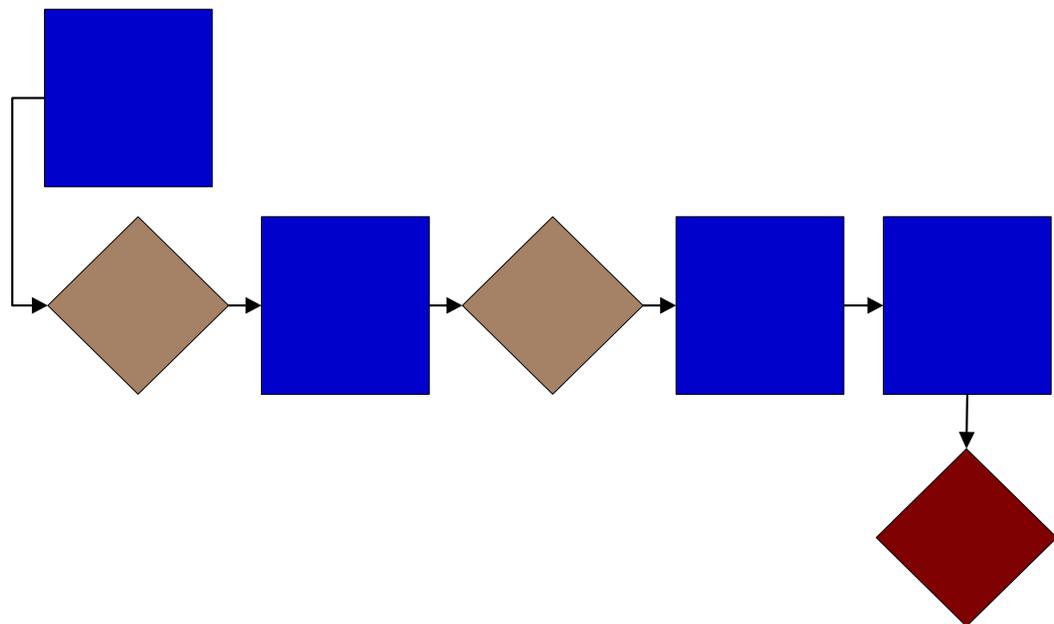
Scoring Quality Procedures

DRC understands the activities and coordination required for data processing and scoring for the GCA Program and has the proven experience and capabilities to

score the tests accurately. DRC brings many years of valuable and accurate scoring experience spanning across programs such as Alabama, Alaska, Louisiana, Pennsylvania, and South Carolina.

We will prepare and refine the requirement documents for the scoring of answer documents well in advance of the receipt of test materials. These specifications will contain detailed scoring procedures, along with the procedures for determining whether a student has attempted a test and whether they should be included in statistics and calculations for computing summary data.

The requirement documents will be completed and reviewed with PDE. After all changes and edits have been made, the final requirement documents will be sent to PDE for final approval. Our process is outlined in the following figure.



Process for Requirements Documents

DRC will ensure all student test booklets/answer documents have been accounted for and processed through scanning, pre-editing, and post-editing processes. Once staff confirms these processes are complete, final scoring processes will begin.

All student answer documents returned to DRC will be scored. The original scanned data will be converted into a master student file. Record counts will be verified against the counts from the Document Processing staff to ensure all students are accounted for in the file. Additionally, a detailed review of the error-tracking log will be performed to ensure any discrepancies are resolved before proceeding with the scoring routines.

DRC’s strict quality procedures can assure PDE accurate scoring. **We are prepared and accustomed to handling programs with multiple forms at various grade levels and/or content areas** and have built-in solid checkpoints

and reviews throughout the entire scoring process. Standard quality inspections will be performed on all data files, including the evaluation of each student data record for correctness and completeness prior to report generation. Student results are kept confidential and secure at all times.

Our Software Quality Assurance staff will ensure the quality of school, district, and state data and make certain that each record is verified for completeness and accuracy. Quality checks will be performed on the data placement and data file formatting for each data element to be displayed on the reports. All data elements will be verified back to the production data file and the data processing rules.

Senior Software Quality Assurance Analysts will conduct a second review to ensure methodology, processes, and procedures are followed and verify that the data files are approved prior to report production.

Score Key Quality

The integrity of item, form data, and score keys will be evaluated in several ways. Similar to our score key validation procedures used on other assessment programs, we will leverage our established, documented process to ensure all score keys are accurate. Test development specialists, psychometric staff, and software quality assurance analysts will check the score keys through a series of validation procedures at varying junctures.

Score Key Quality Procedures

- **Verify for accuracy**—Score keys will be verified for accuracy based on multiple reviews by test development specialists, psychometric staff, and software quality assurance analysts. All item data and score keys will be reviewed and approved by each group prior to scoring GCA Program tests.
- **Take the test**—Multiple staff with specific content knowledge will take each form of the test and compare their results against the score keys on the test maps. The score keys and strand information will again be verified during this step.
- **Score key file import**—DRC will import the approved keys received into our scoring system. Once the keys are successfully imported, software quality assurance staff will re-verify the keys used by the scoring engine.
- **Database accuracy**—All items will be scored in the system using the correct and incorrect item distractors. The database will be validated to make certain the distractor captured in scanning was saved correctly and that the item was given a correct or incorrect answer.
- **Automated system checks**—The scoring engine has automated system checks built-in to validate score keys and proper merging of multiple-choice and open-ended items. Additionally, the software quality assurance team performs independent checks on this data.

Data File Quality Control

DRC understands the critical nature of scoring large-scale assessments. Our systematic approach will ensure successful scoring and 100% accuracy. DRC has

the thorough understanding of the requirements needed to monitor, score, and effectively analyze the data for the GCA Program.

All **data file development for the GCA Program will be done in close association with PDE** to ensure requirements are met. Each data file produced will be **quality checked** for accuracy and completeness a **minimum of three times** by DRC’s Software Quality Assurance Analysts and Project Management staff against PDE-approved layouts, specifications, and processing rules.

Psychometric Quality and Methodology

DRC is the only major testing services provider with a department devoted solely to psychometric quality.

DRC’s Psychometric Services (PS) department is committed to quality and excellence. The department achieves psychometric quality and excellence by assuring that its practices and procedures meet the professional measurement standards outlined in the Standards for Educational and Psychological Testing (AERA, NCME, & APA, 1999). In 2006, **DRC took the initiative and led the industry by starting a department of psychometric quality,**

whose primary focus is the continual improvement of psychometric quality processes, working closely with DRC’s psychometric team. This department, Psychometric Services Quality (PSQ), oversees and ensures psychometric integrity for numerous statewide assessments, including programs for Idaho, Louisiana, Oklahoma, and Pennsylvania. One of the main initiatives of the PSQ department has been the creation of a data forensics system at DRC. This system, called Psychometric Scenes Investigator (PSI), is being designed and developed to systematically and efficiently conduct numerous analyses to ascertain the integrity of test results. PSI will be used and enhanced as necessary to provide data forensic analyses, if desired by PDE. A full description of our data forensics plan is included in *Appendix 8*.

The PSQ staff consists of research analysts skilled in research and measurement theory and methodology. Working closely with our Psychometricians, they will provide psychometric quality control for the GCA Program by:

Psychometrics Quality Procedures

- **Data files**—Quality checks will be performed by this team to verify the integrity of data files.
- **Scored data**—Quality checks will be performed on the data to ensure that test scores have been scored correctly.
- **Calibration, Scaling, and Equating**—Replication of these processes will be performed as an independent quality check.
- **Independent Psychometric Analysis**—Coordinate with the third-party vendors performing independent psychometric analyses.
- **Reports**—Validate that the assessment results are accurate and allow for valid interpretations.

Data File Quality Control

Psychometric quality begins with a check of the student response data file. All fields critical to the analysis, calibration, and equating process are checked and verified by the psychometric and the psychometric quality teams. Variables are validated against the final approved file layout and processing rules to ensure that no unanticipated values exist and that data characteristics appear to be consistent with past experience. All key demographic fields are checked for accuracy. Additionally, a general reasonableness check on the data as a whole is conducted by computing the raw score frequency distribution, verifying the proper numbers of items and the proper location of open-ended items, and by verifying that no unusual or atypical values exist.

Preliminary Item Analysis

Psychometric quality control continues with a preliminary item analysis key check on multiple-choice items. There are many levels of key verification that take place within DRC (i.e., content experts take the exam and compare their keys with the approved scoring key and SQA staff tests and verifies the scoring program), but this preliminary item analysis serves as a final check to identify any items that do not seem to be functioning as expected. The preliminary item analysis is performed on a scored student file as soon as enough student records are available and is performed only on the multiple-choice items. Items that exceed certain criteria in terms of psychometric characteristics are flagged and re-verified by content experts to ensure that the identified item key is correct.

The preliminary item analysis is conducted by our psychometric team and verified by our psychometric quality team. The process is an integral part of ensuring quality and the validity of the test results. The analysis provides assurances that the test is of high quality, and therefore valid inferences can be made from the results.

Calibration, Scaling, and Equating Quality

Check of Calibration Data File

Our psychometric team receives the calibration data file after it has passed all SQA quality control procedures. Once the file is received, the psychometric and psychometric quality teams begin conducting independent quality checks on data fields critical to the analysis, calibration, and equating process. Variables are checked against the final client-approved file layout to ensure that no unanticipated values exist and that other data characteristics appear to be consistent with past experience. All key demographic fields are checked for accuracy. Additionally, a general reasonableness check on the data as a whole is conducted in terms of raw score frequency distributions, proper numbers of items, proper location of OE items, and the presence or absence of unusual or atypical values.

Evaluating the Calibration Sample

After the calibration file has been inspected, our psychometrics staff evaluates the representativeness of the calibration sample. This evaluation includes analyzing key demographic variables (e.g., gender, ethnicity) and comparing them to known state values. Additionally, they examine the p -values of anchor items against their previous operational p -values. These anchor-item p -value plots should demonstrate a reasonably distinctive linear relationship.

Item Analysis

Once the verification of the calibration data files is completed, an additional item analysis is performed on the data by course. The results from this analysis are used as an additional check of the file and re-check of the item key.

Calibration, Scaling, and Equating

The DRC psychometrics and psychometrics quality teams are skilled in both the Rasch and IRT models of measurement. When constructed-response items are included in the assessment, either Master's Partial-Credit Model or the two-parameter generalized partial-credit model (Muraki, 1992) are used. When the Rasch model is used, the psychometrics teams use the WINSTEPS program to run item calibrations. When the two-parameter or three-parameter IRT models are used, the PARSCALE program is used.

The exact calibration, scaling, and equating procedures are specific to any testing program. DRC's psychometricians will work with PDE and their technical advisors to ensure that the chosen procedures are appropriate and meet psychometric standards.

To ensure quality of the calibration, scaling, and equating procedures, all work is conducted independently by a psychometrician and statistical analyst assigned to the GCA Program. Additionally, research analysts from our psychometric quality team independently verify the results.

To ensure comparable scores from year to year, the present year's item calibrations must be linked appropriately to those of prior years. This is achieved through proper implementation of the approved equating steps. To ensure that the equating was done appropriately, a plot of the linking item percent correct scores is constructed (i.e., the current year p -values are plotted against those from the prior year). Similar plots are constructed for the linking item difficulty values. Ideally, these plots should have a strong linear trend. Items straying from the trend line did not perform in the same way both years. Items that depart significantly from the trend line are further evaluated and may be excluded from use in the final equating of items.

Other checks of the equating include verifying and ensuring that the correct anchor items and their correct difficulty values are used, plotting the score distributions for students in each of the achievement levels, and creating longitudinal tables/graphs of scale scores at selected percentiles across years.

These figures and tables provide an important quality check. Small variations across years are expected and may be due to several factors. However, large variations may provide a clue that a systematic error occurred somewhere in the data processing stream.

The results are reviewed by the entire psychometric team assigned to the GCA Program to ensure all results are correct before they are presented to the state and/or its technical advisors.

Independent Third-Party Checks

Calibration and equating are integral parts of the testing program. As an additional verification, DRC will contract with independent psychometric consultants to replicate the entire calibration and equating process, if desired by PDE. The selected third-party vendor undertakes the entire process, starting with item calibration and ending with the creation of the raw-to-scale score tables.

The results from the independent contractors are then compared to those produced by DRC. PSQ staff will work with the consultants to investigate and resolve any discrepancies. Once the results have been approved and verified by DRC, the results are further verified and approved by the state and/or technical advisors.

Student Data Files and Reporting Quality

After the scoring tables are created and verified, the students' raw scores are converted to scale scores based on those look-up tables. The scale scores are applied to the final student data files, individual student reports, and summary reports. All files and reports go through multiple levels of quality checks. The PSQ research analysts serve as one part of that process by performing independent checks on the data files and reports.

Item Bank Process

As previously discussed, all calibrations are run independently by the psychometrician and statistical analyst and subsequently verified by the research analysts. After these independent runs are conducted, the entire team reviews and evaluates the results. Any discrepancies noted, discussed, and resolved before the calibrations are considered final and imported into the item bank system.

Once statistics are put into the item bank, all data elements are checked to ensure that they have been imported without error. Lastly, sample data cards are printed from the system and checked to ensure that proper statistics and values are displayed correctly.

Quality: A Corporate-Wide Value at DRC

As described above, DRC has in place the necessary quality control processes—from initial project planning through the delivery of final reports—to successfully develop and administer large-scale assessment programs. Through our Quality Management System, DRC feels confident in guaranteeing the accuracy and on-

time delivery of our large-scale assessment projects. We look forward to providing these high-quality services to PDE.

VII.I.4. Creation of Data Files

Student responses to multiple-choice items, as well as demographic information, will be captured as images and preserved for use during the image scoring process. Information embedded in the student precode label or the district/school label will also be captured during scanning. This information will link back to the PIMS student record or to the site at which the student tested (if a school/district label was used). Booklet counts and page integrity will be maintained throughout the scanning process by storing data in a Relational Database Management System (RDMS) using unique identifiers that link each image to a single, individual record, preserving school/district and other identification and demographic information. A relational database significantly increases system flexibility and provides for robust data analysis capabilities.

Once the demographic information and multiple-choice data pass all the pre-defined editing processes, the images of the student responses to open-ended (OE) items are extracted into files for scoring. The OE student response images will be routed through the DRC Imaging Workflow System to handscoring terminals at DRC's Scoring Centers for scoring by qualified readers. Images are stored so that they can be efficiently retrieved on the basis of student and school identification information, scores, and item information. Upon completion of processing, scannable documents are boxed for security purposes and final storage. Please see *Subheading VII.I,16, Storage, Retrieval, and Destruction of Materials*, for more information regarding our storage and retrieval procedures.

Data File Accuracy

Below, we present an overview of the processes and methods DRC will use to ensure the accuracy and completeness of student data and associated data files:

- All student response documents returned to DRC will be scored. Multiple-choice items and demographic information will be image scanned and the original scanned data will be converted into a master student file. Open-ended items will also be image scanned and sent electronically to DRC Scoring Centers for scoring by trained, qualified readers. All student information and score results are kept secure and confidential throughout the scanning, scoring, and reporting processes.
 - All student answer documents will be reported. No invalidation process was requested in the RFP or Questions and Answers. Should PDE desire this service, pricing will be provided upon award.
- Record counts will be verified against the counts from the Document Processing staff to ensure all students are accounted for in the file. Additionally, a detailed review of the error-tracking log will be performed to ensure any discrepancies are resolved before proceeding with the scoring routines.

- The scoring process will include the scoring of multiple-choice items to the answer key and the aggregation of raw scores from the open-ended responses. Using the raw scores, scaled scores will be calculated.
 - DRC will ensure the accuracy of the answer keys. At least two DRC staff with content-specific expertise will take the test and compare their answers to the answer keys on file for each test item. Test-takers will also verify academic standard information. Quality Assurance staff will compare the processing file against the answer key source file to ensure accuracy.
 - Once the multiple-choice keys have been analyzed and approved for accuracy, clean, edited batches will be processed through scoring and reporting programs. Scoring programs will contain answer keys and academic standards categorizations for each item. Items will be scored as right, wrong, omitted, or double-gridded. After scanning and scoring, DRC Quality Assurance staff will verify values for multiple-choice and open-ended items.
- After scanning and scoring, DRC's Software Quality Assurance will perform an item response frequency analysis on both initial and complete data sets. DRC Psychometrics staff will perform item response frequency analysis, independent foil analysis, and differential item functioning analysis (please see *Subheading VII.J., Analyses of Assessment Data*, for a thorough discussion of DRC's proposed data analysis procedures).
- Raw-to-scale score conversion tables and cut points based on pre-equating are provided to DRC Information Systems Team. The conversion table allows each student's raw score to be converted into a scaled score. The cut points are used to assign each student to a proficiency category.
- Additional reporting software will contain the procedures for sorting and summarizing data. Our Software Quality Assurance staff will ensure the quality of school-, district-, and state-level data and make certain that each record is verified for completeness and accuracy. Quality checks will be performed on the data placement and data file formatting for each data element to be displayed on the reports. All data elements will be verified back to the production data file and the data processing rules. Senior Software Quality Assurance Analysts will conduct another review to ensure methodology, processes, and procedures are followed and verify that the data files are approved prior to report production.
- All data files for the GCA Program reports will be quality checked for accuracy and completeness by DRC Quality Assurance Analysts.
- All data file design, development, and enhancement efforts will be done in close association with PDE to ensure that all requirements for reporting are met.

VII.I.5. Editing Procedures for Scanning

After scanning, the documents are processed through a computer-based editing program to detect potential errors in specified response fields. Pre-defined edit specifications will be mutually developed by DRC and PDE. Marks or omits that do not meet the pre-defined editing standards are flagged and routed to the Document Processing editing staff for resolution.

Using unique serial numbers printed on the documents during scanning, the editors compare the actual documents to online data. Corrections are then made to the data file according to pre-defined, program-specific guidelines. The editing staff follows strict quality-control procedures to produce clean data files that can be submitted for scoring and reporting functions.

Quality-Control Procedures for Editing

- Edit specifications will be developed mutually by PDE and DRC.
- Experienced editors meticulously review any potential irregularities detected during scanning and make necessary corrections online to the image data file, referring to the actual document as required.
- Editors determine if the marks are valid (based on assessment requirements) or non-correctable.
- All items flagged during document scanning are presented to editing staff for first-time entry.
- Any changes made to scanned values and all items entered the first time are double-keyed for verification by editing quality staff.
- Once verification by editing quality staff is completed, a quality-control report is generated for review during post-editing.
- After all corrections for a batch have been entered and verified, the correction file is stored in a relational database for reference.

VII.I.6. Pre-Editing Procedures

Quality-control procedures are critical to DRC's document scanning process. All image scanning programs will go through quality review before testing materials arrive at our facilities. Throughout the scanning process, batches will be checked for quality and scanning accuracy by experienced Document Processing staff. All scanners are calibrated and cleaned on a regularly scheduled basis to ensure accurate and consistent scoring. DRC also has on-site field service engineers to resolve any technical issues as they arise.

DRC's scanning process produces comprehensive, detailed information, including:

- Student demographic data.
- Student multiple-choice response data.

- TIFF images of complete documents.
- Identifiers to link the TIFF images to the student demographic data.

Our quality-control procedures for document scanning and pre-editing are highlighted below.

Quality-Control Procedures for Document Scanning and Pre-Editing

- Scope of Work Agreements (SOWA) will be established, with approval by PDE if desired. All processing and scanning staff will adhere to the requirements contained in the SOWA.
- As scanning occurs, a unique serial number is printed on each sheet of paper. This serial number ties documents together and maintains sequencing within batches.
- The scanners pick up pre-defined processing criteria related to pre-printed barcodes, multiple-choice items, and student demographic and identification information. Open-ended item images are scanned and separated out for image-based handscoring.
- Regularly scheduled calibration and scanner cleaning processes will be conducted to ensure image and read-level quality and consistency.
- As documents are scanned, the scanner is monitored to ensure that images meet DRC's strict quality standards.
- The Document Processing Supervisor will conduct a review of the entire first batch prior to full production to ensure error-free processing.
- The Document Processing Supervisor will conduct daily reviews of quality reports for each editor to monitor the accuracy of the editing process.
- All scanned images will pass through a software clean-up program and process to detect potential issues (double marks, smudge marks, omits, etc.).
- After image clean-up, a random sample of images is presented for image quality approval. If any image fails to meet DRC's quality standards, the entire batch of documents is rescanned.
- Page-scan verification is performed to ensure that all pre-defined portions of a document were correctly captured. A flatbed scanner is used to capture responses and images for any missing pages. These images are then added to the image data file and merged with the appropriate document.

VII.I.7. Post-Editing Procedures

After all scanning and editing processes have been completed, and before images are released for scoring and student-level processing, a final edit is performed to confirm that all requirements for final processing have been accurately met.

Quality-Control Procedures for Post-Editing

- During this processing step, the actual number of documents scanned are compared to the number of scannable documents assigned to the box during Book Receipt. Any count discrepancies between Book Receipt and documents scanned will be resolved at this time.
- Suspect student precodes, district and school numbers, and documents IDs are reviewed for additional verification.
- All editing quality-control reports are reviewed to ensure all changes were processed accurately.
- All corrections during post-editing are made electronically and a new validation report is generated to confirm the changes have been processed accurately and the report is clean.
- After all final processing requirements have been met, the batch is released for OE scoring and student-level processing.

VII.I.8. Plan for Standards Setting for all GCAs

DRC's proposed plan for standard setting for all GCAs is included under *Subheading VII.J.8.*

VII.I.9. Scoring Multiple-Choice Items

Our 16 years of experience providing scoring for Pennsylvania's testing programs gives DRC **a unique understanding of the activities and coordination required for data processing and scoring of the GCA Program.** Additionally, DRC scores over 5 million student answer documents on an annual basis for numerous assessment programs and has successfully scored and matched multiple-choice and open-ended responses without any reported errors.

Delivering assessment results on time and without error is critical. Our scoring and reporting systems have quality procedures integrated throughout, including both automated and manual inspections, to ensure data accuracy. DRC's experience and expertise will directly contribute to the successful processing and reporting for the GCA Program within the prescribed time limits.

The multiple-choice items will be scored against the appropriate answer key, indicating correct and incorrect responses. In addition, the student's original response string will be stored for data verification and auditing purposes. We will prepare and refine the requirements documents for the scoring of answer documents well in advance of the receipt of test materials. These specifications will contain detailed scoring procedures, along with the procedures for

determining whether a student has attempted a test and whether the student should be included in statistics and calculations for computing summary data. DRC will ensure that all answer keys have been approved and verified for accuracy prior to the scoring of any student responses. Student scale scores and performance levels will be determined prior to the production of final data files and reports.

Student responses to open-ended items will be scored at DRC Scoring Centers by trained, qualified readers. Scores for a student's responses to open-ended items will be systematically matched to that student's multiple-choice data by a unique document ID (lithocode). This process allows DRC to create a **single, accurate, reliable data record for each student assessed** by linking all score and demographic data for a specific student, including precode data and scores collected during scoring of multiple-choice and open-ended items.

DRC's strict quality procedures will result in accurate scoring. We are prepared and **accustomed to handling programs with multiple forms and assessments** at each grade level and have built-in solid check-points and reviews throughout the entire scoring process. Our process and established quality-control procedures have repeatedly proven the accuracy of our systems.

Once the scored master student file is deemed 100% accurate, DRC's psychometrics staff will perform additional detailed analysis on the data files prior to PDE's review and approval process.

Standard quality inspections will be performed on all data files, including the evaluation of each student data record for correctness and completeness. Student results will be kept confidential and secure at all times. Please see above for more information on DRC's Quality Assurance procedures for scanning and scoring computer/software programs and data accuracy. Please refer to *Subheading VII.K., Reports and Data Files*, for additional information about the creation of data files.

Software Quality Assurance Testing for Data Analysis and Verification

To provide PDE with the highest level of accurate test results, **DRC will conduct a thorough evaluation of all scored data**. File formats and data elements will be validated against client-approved layouts, specifications, and processing requirements. Detailed test scripts will be executed to confirm accuracy. Some of the steps include quality verification of:

- Answer keys/test maps
- Raw scores
- Raw-to-scale score conversions
- Scale-score comparisons to performance achievement levels

- Disaggregated data
- Processing rules for individual student and summary level data

The quality assurance steps involve processing sample student records through the data processing and scoring system. Each student's data record will be carefully reviewed and evaluated to ensure it was **scored with 100% accuracy**.

To reduce the risk of human error, our Software Quality Assurance programmatic test routines will be used to thoroughly evaluate each student's data record that will be produced for use in final data files and reports.

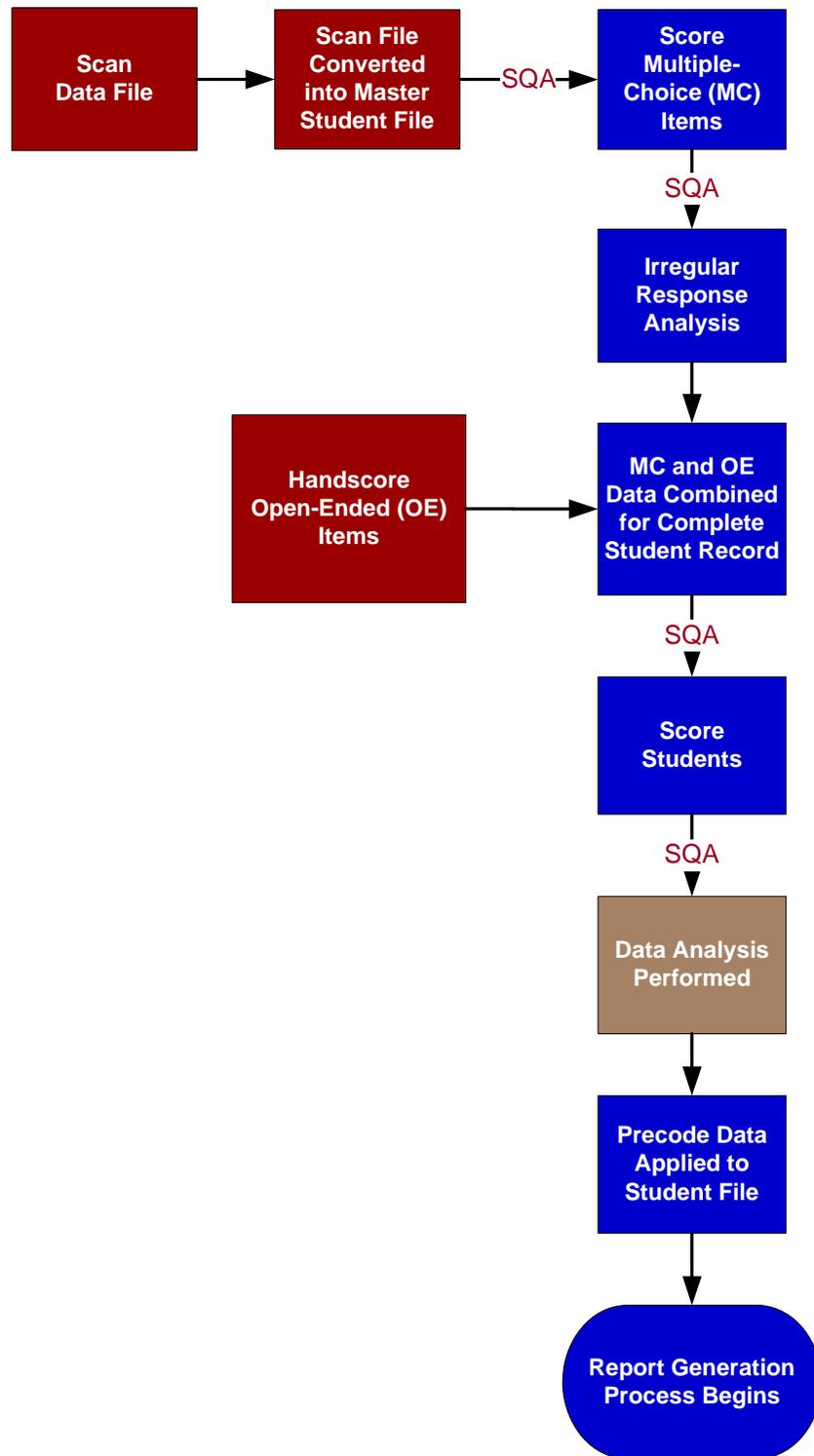
Score Key Quality

The integrity of item, form data, and score keys will be evaluated in several ways. We will leverage our established, documented score key validation process to ensure all score keys are accurate. Test development specialists, psychometric staff, and software quality assurance analysts will check the score keys through a series of validation procedures at varying junctures.

Score Key Quality Procedures

- **Verify for accuracy**—Score keys will be verified for accuracy based on multiple reviews by test development specialists, psychometric staff, and software quality assurance analysts. All item data and score keys will be reviewed and approved by each group prior to scoring GCA Program tests.
- **Take the test**—Multiple staff with specific content knowledge will take each form of the test and compare their results against the score keys on the test maps. The score keys and strand information will again be verified during this step.
- **Score key file import**—DRC will import the approved keys into our scoring system. Once the keys are successfully imported, software quality assurance staff will re-verify the keys used by the scoring engine.
- **Database accuracy**—All items will be scored in the system using the correct and incorrect item distractors. The database will be validated to make certain the distractor captured in scanning was saved correctly and that the item was given a correct or incorrect answer.
- **Automated system checks**—The scoring engine has automated system checks built in to validate score keys and proper merging of multiple-choice and open-ended/writing-response items. Additionally, the software quality assurance team performs independent checks on this data.

The figure below outlines DRC's multiple-choice and open-ended scoring and quality process.



Process for Merging Student Data and Scores

VII.I.10. Scoring Open-Ended Items and Training Procedures

DRC brings a tremendous amount of experience scoring student open-ended items for Pennsylvania, and we have total confidence in our ability to continue doing a superior job with the proposed Graduation Competency Assessments. DRC looks forward to offering the same high quality performance as we have for the PSSA for the past 16 years. We have 23 years of experience in successfully providing accurate scores for millions of student responses in numerous content areas, including mathematics, reading, writing, science and social studies. Our statewide assessment experience is summarized below:

- **Mathematics**—Alabama, Alaska, California, Louisiana, New Jersey, North Carolina, Ohio, **Pennsylvania**, South Carolina, and Washington
- **English Language Arts**—Alabama, Alaska, the International Reading Assessment, Louisiana, New Jersey, North Carolina, Ohio, Oklahoma, **Pennsylvania**, and South Carolina
- **Science**—Alaska, Louisiana, Ohio, **Pennsylvania**, and South Carolina
- **Social Studies**—Louisiana and Ohio
- **Graduation Assessments**—Alabama, Alaska, Louisiana, and Ohio

DRC will continue to work closely with PDE to ensure that Pennsylvania students' responses are assessed using the scoring guidelines and anchor sets that will be developed in collaboration with PDE and Pennsylvania teachers. A minimum of 10% of all reading, math, science and social studies responses will be scored independently by two scorers for the purpose of reliability. All English compositions will be scored twice by two independent readers. If a score is non-adjacent, a third, expert reader will score the response to resolution.

Our Performance Assessment Services (PAS) content specialists have multiple years of experience working with the PSSA, and our management staff has firsthand knowledge of monitoring Pennsylvania projects, producing accurate reports, and meeting deadlines by working with the Commonwealth since 1992. We understand that the deadlines to meet graduation dates will be even more stringent for the GCA and we are confident that we can continue to perform to PDE's expectations.

We are very proud of the fact that we will be able to continue utilizing key personnel for each content area. For example, we have had the same people attend rangefinding for the PSSA over the past eight years and this is invaluable when it comes to maintaining Pennsylvania scoring standards and ensuring consistency from one year to the next. This continuity has been one of the central reasons that DRC's handscoring staff has been able to meet stringent deadlines with accurate results and will continue to do so for the GCA.

Our references will confirm that DRC consistently meets our handscoring deadlines and maintains our focus on quality throughout the handscoring process. We believe that our experienced personnel, quality training materials, and thorough quality-control measures are all essential to the success of a handscoring project. We would welcome the opportunity to provide PDE with this excellent service under the new GCA contract.

Handscoring Staff

DRC offers experienced personnel who are unsurpassed in the industry. More specifically, all of our proposed PAS staff members have worked on the PSSA for many years. In fact, Ms. Sue Drexler, DRC's proposed Handscoring Manager, has been working with Pennsylvania's assessment programs since 1994. Her experience with rangefinding and managing all aspects of handscoring is invaluable moving into the GCA Program.

Our team of scoring and content specialists consists of leaders in the handscoring field; many of them facilitate professional development sessions during the off season. Our combined experience includes designing and monitoring the scoring of many large-scale assessment efforts that have incorporated numerous content areas, scoring models, and procedures. DRC is proud of its reputation of working diligently with clients to customize scoring to meet the specific needs of a particular assessment.

DRC is confident that our Scoring Directors will continue to adhere to Pennsylvania's scoring standards; we truly have an in-depth understanding of the standards being assessed. All of our Scoring Directors will be on-site throughout scorer training and scoring. Our handscoring team welcomes all input and requirements from PDE and believes that strong communication between vendor and client is necessary for any successful assessment program.

Staff Profiles

Below we present an overview of each of our proposed PAS team members. Complete profiles are provided in *Section E, Personnel*, and résumés are provided in *Appendix 1*.

Ms. Sue Drexler, Director, Performance Assessment Services, is proposed as the **Scoring Manager** for the GCA Program. Ms. Drexler has over 18 years of experience working on statewide assessments. She has served as the Director of Performance Assessment Services for the PSSA since 1994. In addition to her work on the Pennsylvania assessments, she has worked on several other statewide testing programs and has served as a manager for several. Her DRC experience also includes item development and professional development. She has been with DRC's Performance Assessment Services since 1990.

Mr. Warren Hite, Content Specialist, is our proposed **Algebra 1, Algebra 2, and Geometry Handscoring Lead** for the GCA Program. Mr. Hite has 15 years of experience working on large-scale assessments, with a focus on the content

area of mathematics. Currently, he is involved almost exclusively in assessment activities for Pennsylvania. In addition to the PSSA, Mr. Hite has worked on numerous other statewide assessments. He has been a part of the DRC team since 1993.

Mr. Nick Hook, Content Specialist, is DRC's proposed **Literature Handscoring Lead** for the GCA Program. Mr. Hook has 16 years of experience working on large-scale assessment projects, having begun his work as a DRC scorer in 1992. Currently, Mr. Hook works exclusively on handscoring services for the PSSA, a program on which he has worked since 1995. He specializes in the content areas of reading and English language arts and works closely with DRC's Test Development Department to proofread and edit rubrics for clarity in training. He has also been involved with other handscoring projects for several other states.

Ms. Wendy Marik, Content Specialist, Performance Assessment Services, is DRC's proposed **English Composition Handscoring Lead** for the GCA Program. She has been involved in the handscoring of PSSA writing assessments for the past eight years. Ms. Marik has a total of 10 years of experience working on large-scale assessment projects, having joined DRC in 1998. Ms. Marik has experience with numerous writing assessment models and rubrics. She has worked with clients to implement transition from domain to holistic scoring guidelines. During her time at DRC, she has applied her expertise to many statewide assessments in a variety of subject areas, such as writing, mathematics, English language arts, social studies, science, and practical living skills. She has worked on testing programs for numerous other DRC state clients.

Ms. Violeta Lee, Science Content Specialist, is proposed as the **Biology and Chemistry Handscoring Lead** for the GCA Program. Ms. Lee has more than 10 years of experience in science assessment design, development, evaluation, and scoring and more than eight years of direct PSSA handscoring experience, including four years as the Mathematics Lead for the Spanish translation version of the PSSA. In her role at DRC, she specializes in the development of training materials for readers and training leaders, analytical and holistic scoring rubrics, selection of exemplar item sets for training and rangefinding applications, and the facilitation of rangefinding activities. Ms. Lee has been with DRC since 1997.

Mr. Bob Dzandzara, Social Studies Content Specialist, is proposed as the **World History, United States History, and Civics and Government Handscoring Lead**. Mr. Dzandzara has 19 years of experience working on large-scale assessment projects. He has worked with a variety of clients on social studies assessments since 1995. His responsibilities include working with state department personnel and educators to develop training materials, monitoring handscoring projects and supervising handscoring staff, and coordinating rangefinding activities. During his 19 years with DRC, Mr. Dzandzara has worked on numerous assessment projects including social studies, writing, and English language arts.

Ms. Alison Lyder, Vice President, Performance Assessment Services will serve as a **Senior Handscoring Advisor** for the GCA program. She will provide her expertise and assistance to the Scoring Manager and Content Area Handscoring Specialists. She has extensive Pennsylvania-specific program experience, having worked on PSSA scoring projects for the past 12 years. Throughout her career, Ms. Lyder has worked on large-scale assessment projects in numerous capacities. Her 26 years in the assessment field, including 21 years with DRC, give her an in-depth understanding of the kinds of information that must pass between the client and the scoring staff to ensure that student scores reflect the intentions of the assessment.

DRC Scoring Facilities

Our Minnesota scoring facilities are located in Minnetonka, Woodbury, and Brooklyn Park, all in close proximity to our corporate headquarters in Maple Grove. Additionally, we operate three other scoring sites: one in Sharonville, Ohio, a suburb of Cincinnati; one in Columbus, Ohio; and one in Austin, Texas. **We have maintained a handscoring site in Harrisburg, Pennsylvania, since 2004 and will open a scoring facility in the Pittsburgh area in February 2009.**

Each site has ample square footage that is divided into several large, open scoring rooms. Additionally, there are numerous small offices that are used for conferences and small-group training. Each site maintains a large break room equipped with coffee, vending machines, and refrigerators, as well as additional amenities.

Scorer workstations are designed for comfort and ease of use. Large, flat-screen monitors display crisp images and reduce eyestrain and glare. Scorers can view an entire student response page without scrolling. Ergonomically-correct, adjustable chairs, tables, and keyboard/mouse trays keep our scorers comfortable and focused. Two imaging stations sit on each sturdy table. DRC management makes certain that the reading rooms are kept very quiet during scoring. The scoring facilities are accessible to the physically challenged and are convenient to major highways and airports.

At each of our sites, the DRC Scoring Facility Technical Coordinator resolves any technical issues that may arise. Additionally, a Scoring Site Coordinator and a Human Resource Coordinator attend to human resource and facility management needs. Together with the Scoring Project Director, these scoring site personnel form a support team dedicated to maintaining day-to-day operations and allowing the Scoring Directors to focus exclusively on maintaining consistently accurate scoring.

Handscoring Security

Each Scoring Center is a secure facility. Access to each facility is limited to staff and to visitors who are accompanied by authorized staff. All scorers are made aware that no scoring materials may leave the Scoring Center and must sign legally binding confidentiality agreements before work begins. To prevent the

unauthorized duplication of secured materials, scorers are not able to print from their imaging stations. Scorers only have access to student responses they are qualified to score. No demographic data is available to scorers at any time.

For the image scoring process, each handscoring site is connected to the DRC main operation facility with multiple T1 transmission lines. The main operation facility has secure database servers and multiple applications that support the handscoring processes. Daily tape backups are done for production databases and images, with tapes being rotated off-site for disaster recovery purposes.

DRC strictly maintains the security and confidentiality of all items and student responses. No secure materials will be released to rangefinding committee members, and no materials may be removed from DRC facilities. As part of the introductory orientation and training, DRC will discuss Pennsylvania security guidelines with committee members during the opening portion of each session and obtain signed security agreements from all rangefinding committee members before materials are distributed. DRC will retain these agreements for the duration of the contract.

Image Handscoring

Now in its seventh year of operation, DRC's Image Handscoring System has proven to be highly efficient and completely accurate for scoring the PSSA. Our dynamic system allows scorers to score items online, increasing efficiency by eliminating the routing of paper and eliminating the possibility of lost student answer documents. Instead, imaged responses are electronically routed to geographically dispersed DRC Scoring Centers. Responses are allocated to scorers through a custom dealer program, ensuring that each scorer is assigned a random workload that allows the project to be processed in the most efficient manner.

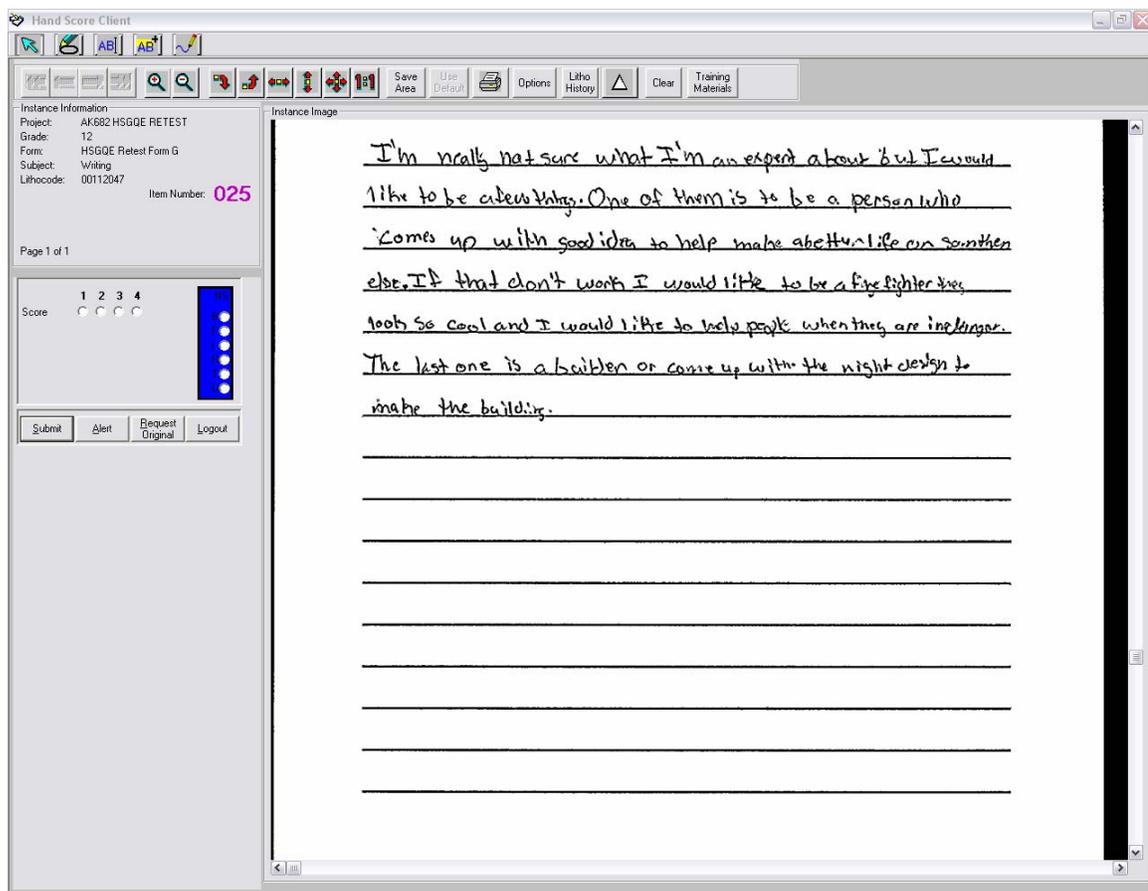


With the Image Scoring System, each open-ended item's position within the student answer document is defined through the use of the programmatic Item Definition Application. For each item, the system also requires the definition of the possible score values, the possible non-score values, the applicable scoring

rule (e.g., 10% of responses are read twice), and the Scoring Center at which the item will be read.

The Image Handscoring functionality also requires scorers to forward all non-scorable responses to the Scoring Director. Only the Scoring Director is able to assign the non-scorable code. Scorers will also “alert” any responses that indicate potential issues related to the student’s safety. The alert process is described more fully later in this subheading.

The system provides the scorers with the ability to view full-page images from multiple perspectives, such as zoomed in/out and image flipping or rotation, to correctly interpret written responses. Images remain intact with the various viewing capabilities and cannot be modified by the scorers. Additionally, the Image Handscoring functionality applies a set of process rules and client-defined read-behind criteria.



Example of Student Response Display for a Writing Assessment

Each handscoring site is connected to the main DRC operations facility with multiple T1 transmission lines. The operations facility has secure database servers and multiple applications that support the scanning, editing, scoring, and handscoring processes. Database backups and archived images are stored off-site

on tape media for disaster recovery purposes. Each DRC scoring site has a server and a local area network (LAN). Scorers, Team Leaders, and Scoring Directors connect to the LAN via hundreds of PC workstations and use locally resident software to view and score student responses. Authorized on-site DRC personnel (e.g., Content Specialists, Project Managers) can access the LAN to recall images of any student document.

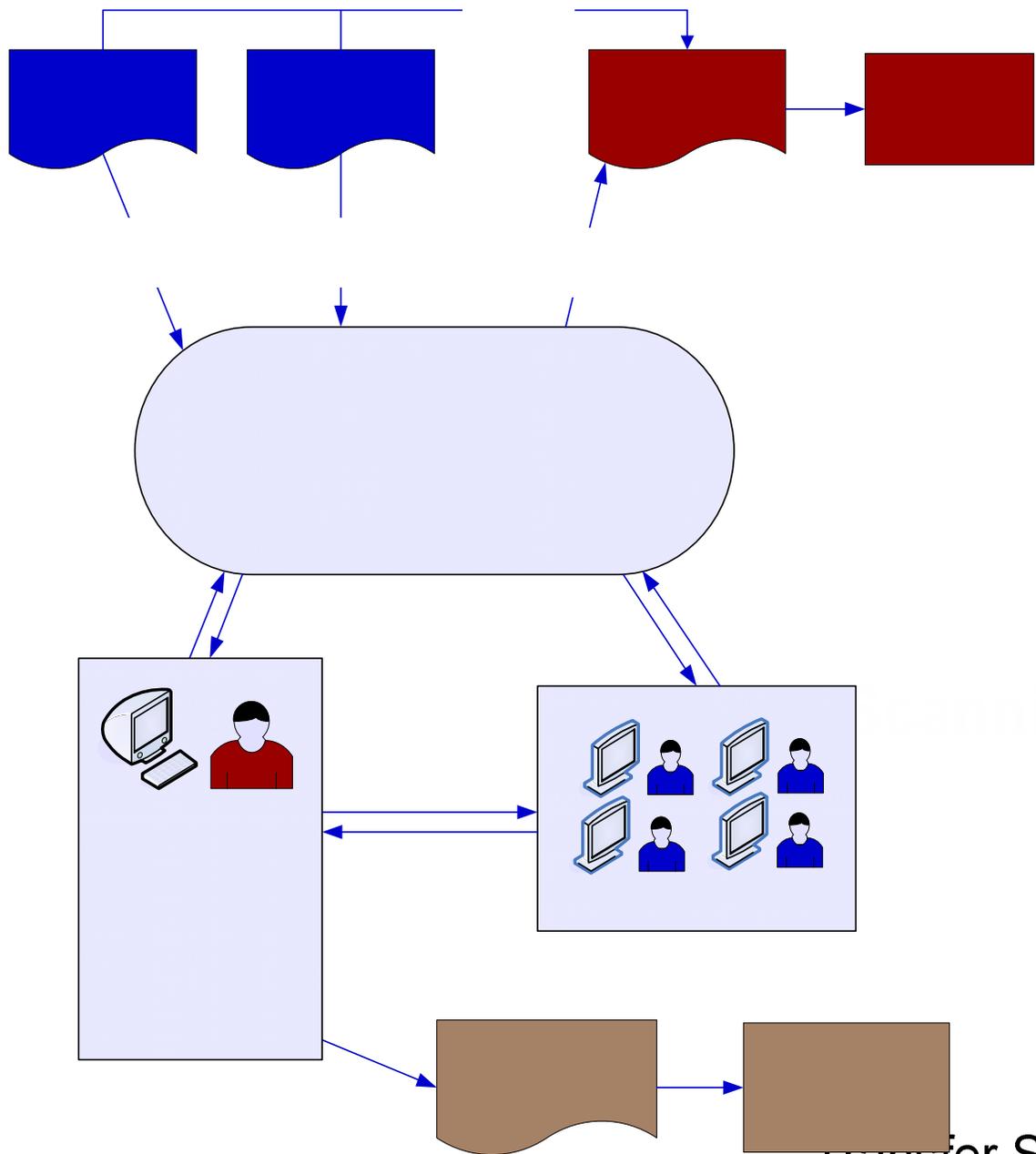
Quality Control of the Handscoring Imaging System

Software Quality Assurance Analysts test the imaging system to verify that all handscoring programs are compliant and in place for performance assessment personnel prior to the transfer of production images. Images produced from test scan files are randomly distributed to handscoring computer terminals, where Quality Assurance Analysts score the test images using the Pennsylvania handscoring criteria and specifications. Throughout this testing cycle, multiple quality checks are executed to ensure that the data integrity for each student record is intact and accurately reflected in the scoring database.

The handscoring quality-control reports are also tested by Quality Assurance Analysts during this process to ensure that the performance assessment personnel will be able to track scorer reliability, score point distribution, and item status throughout the handscoring phase of the GCA Program.

Additionally, the Image Handscoring functionality applies a set of process-rules and client-defined “read-behind” criteria. This functionality includes a “50% rule” that ensures that one scorer cannot score more than 50% of a student’s responses in any given content area, which helps to disperse responses across multiple scorers. This quality-control measure ensures that multiple scorers contribute to each student’s final results, thereby producing results that are more psychometrically sound.





DRC Image Handscoring System and Process

Transfer Student
Response Images

Rangefinding

DRC understands that rangefinding meetings will be held in Pennsylvania and that we will be responsible for travel, food, and lodging expenses. Meetings for mathematics, reading, science, and social studies and writing will include five teachers per subject. DRC will be responsible for collecting information to award Act 48 hours to those who attend. DRC will also investigate possibilities for teachers to receive other attendance incentives such as gift cards for office supplies or books, and possibly the opportunity to participate in scoring at the

Harrisburg Scoring Center. DRC would like to discuss options for incentives with PDE upon award.

DRC will coordinate rangefinding meetings for the GCA to select the training responses necessary for scoring. The course-specific committees for these meetings will be composed of PDE staff, PAS content specialists and Scoring Directors, and Pennsylvania educators. Committee members will be selected from a pool of applicants from across the Commonwealth. DRC fully understands that PDE has final approval on the list of invited participants. These committees of educators will review and score samples of student work and reach an agreement about the application of scoring guidelines. DRC will then take these scored responses and create handscoring training material for each item and prompt.

DRC content specialists and Scoring Directors will prepare for rangefinding meetings by using our Imaging Rangefinding Viewer to access student responses. They will use DRC/PDE developed scoring guidelines for reading, mathematics, science, social studies, and writing to select a representative sampling for each score point. These responses will be assembled into sample sets and duplicated for use at rangefinding.

DRC and PDE staff will begin the meetings by reviewing specific student responses to ensure that committee members are anchored to the same parameters. When an understanding of the scoring guidelines has been established, committee members will score the responses independently and then the scores will be discussed until a consensus is reached. Facilitators will move from item to item/prompt to prompt until the committee members have scored a sufficient number of responses to construct training, qualifying, and recalibration sets. Only responses with a high level of agreement will be used to train our scorers. DRC staff will make careful notes of scoring decisions made by the committees and PDE content staff for use in training the scorers.

As timelines have become more challenging in the industry, DRC has adapted the process of assembling training materials that were reviewed while conducting thorough field test rangefinding. From these meetings, we come away with solid anchor responses and abbreviated training materials. Once field test items become operational, DRC then uses those anchors established at field test rangefinding to supplement further materials for that particular item if it becomes operational. Our content specialists have found this to be an accurate, efficient and cost saving process. This process has yielded quality PSSA training materials in the past, and we are confident it will be successful in the future.

Training the Scorers

DRC selects scorers who are articulate, concerned with the task at hand, and, most importantly, flexible. Our scorers must have strong content-specific backgrounds: they are educators, writers, editors, and other professionals. They are valued for their experience but, at the same time, are required to set aside their own biases

about student performance and accept the scoring standards of the client's program.

With the excellent and well-educated labor pools available near all our scoring facilities (stemming, in part, from the numerous colleges and universities in the areas), DRC is able to tailor its scorer staff to the client's program. All of our scorers for the GCA will have a four-year degree with a background in the content areas being assessed, and a demonstrated ability to write.

DRC has a Human Resources Coordinator dedicated solely to recruiting and retaining our scorer staff. Applications for scorer positions are screened by the Project Director, the Human Resources Coordinator, and recruiting staff to create a large pool of potential scorers. In the screening process, preference is given to candidates with previous experience scoring large-scale assessments and with degrees emphasizing the appropriate content areas. At the personal interview, scorer candidates are asked to respond to a DRC writing topic. Additionally, candidates must demonstrate proficiency in the content areas they will be scoring. For example, mathematics scorer candidates must successfully solve a series of DRC mathematics problems and show all steps necessary to reach the correct answer.

For all assessments scored in DRC Scoring Centers, DRC will provide Team Leaders who will assist the Scoring Directors with scorer training and monitoring. Comprehensive Team Leader training lasts approximately two days. The Scoring Director for each content area will direct this training. Team Leader training will follow the procedures that are used in the scorer training (detailed below), but will be more comprehensive due to the training responsibilities required of the Team Leaders. During their training, Team Leaders will be required to annotate all of their training responses with official PDE/DRC annotations. To promote room-wide scoring consistency, it is imperative that each Team Leader imparts the same rationale for each score assigned. Training by the Team Leaders should result in each scorer assigning correct scores for the correct reasons. Once the Team Leaders have qualified, they will prepare for the arrival of their teams of scorers. Teams will consist of 10 scorers per team.

Training of scorers for the GCA will begin with a room-wide presentation and discussion of the scoring guide by the Scoring Director. Next, the scorers will practice by scoring the responses in the training sets. Afterwards, the Scoring Director and/or Team Leaders will lead a thorough discussion of each set.

After the scoring guide and all training sets have been discussed, scorers must demonstrate their ability to apply the scoring criteria by qualifying (i.e., scoring with acceptable agreement with true scores) on at least one of the qualifying sets. Any scorer who does not qualify by the end of the qualifying process will not be allowed to score actual student work from the GCA Program.

DRC insists that scorers work the full core hours. Our core hours are 8:30 to 4:00. Additionally, scorers must be present for each training session or they cannot

score that particular item. This has always been a steadfast policy at DRC. DRC will take all necessary steps to anticipate the correct amount of time needed to score each item. We realize that staying on schedule is of the utmost importance in order for student results to be reported in the timeframes required by the RFP.

Training and Qualifying Materials Procedures

DRC's Scoring Directors will use the collection of scored responses from rangefinding to compile the materials needed to train scorers. Responses that are particularly relevant (in terms of the scoring concepts they illustrate) will be annotated for use in the scoring guide as anchor papers. Approximately three anchor papers will represent each score point in the scoring guide. The scoring guide for each item will serve as the scorers' constant reference. For each item being scored, we typically develop two to three training sets and two qualifying sets of 10 responses per set. More training/qualifying sets can be developed for particularly complex items. DRC currently requires 70% exact agreement as the qualifying standard for mathematics, reading science, social studies and writing for the PSSA and recommends the same standard for the GCA Program.

Recalibration sets will be administered on a regular schedule to ensure that the scorers are consistently scoring with accuracy throughout the project and to provide for retraining of scorers on a regular basis.

In addition to training, qualifying, and recalibration sets, DRC will choose responses for use as validity sets. The distribution of the validity papers will be high at the beginning of the scoring window and will decrease as agreement levels are met.

A fundamental objective of any handscoring activity is that results be accurate and consistent. Therefore, it is important that consistently high-quality methods of training and monitoring scorers are developed and employed consistently with all assessments.

Scoring Procedures

Over the past 16 years, DRC has worked with PDE to refine our scoring procedures. We believe that our training, scoring, qualifying, and monitoring processes are the best in the industry. All of these processes have been used for years to score the PSSA, and we will uphold the same level of dedication to accuracy and quality upon award of the GCA contract:

- Pairs of scorers will be seated in ergonomically adjustable chairs at long, rectangular tables. There are two imaging stations at each table. Each workstation will include a large flat-screen monitor for clean image reproduction and easy viewing. Each scorer will be assigned a unique ID number and password.
- The Scoring Director will explain in detail the directions for use of the computerized handscoring system. All scorers will review the Imaging Handbook, created specifically for DRC scorers.

- The student responses are separated for scorers by item and subject. Images of specific sets of items (module-specific) will be sent randomly to designated groups of scorers qualified to score those items. Only qualified scorers will have access to student response images. The scorers will read each response and enter the correct score. After the score is entered, a new response image will appear.
- A minimum of 10% of all responses will receive a second independent reading for math, reading, science, and social studies. Writing compositions will receive two independent readings with adjudication of non-adjacent scores. Our imaging system ensures that all responses are properly routed to two separate scorers who are qualified to score the item. Scorers do not know if they are reading a response for the first or second time: all first and second readings are “blind.”
- Ongoing quality-control checks and procedures will monitor and maintain the quality of the scoring sessions. If any unusual data are observed, DRC will investigate and resolve any issues.
- Routing and scoring sets of imaged items continues until all items or prompts have received the prescribed level of first and second readings.
- DRC’s Image Scoring System allows for on-demand retrieval of specified images (e.g., specific batch files, specific courses, specific students) should the need arise during or subsequent to the handscoring process.

Handscoring Quality Control

Accurate and consistent results are the backbone of all handscoring activities. The following methods used by DRC guarantee scoring quality:

- **Anchors** (or benchmarks) are pre-scored student responses used to define and exemplify the score scale. For each score point, anchors will be selected to reflect the entire range of performance represented by that score based on the judgment of the rangefinding committee and PDE/DRC Content Specialists. The anchors, which will be included in the scoring guide and training sets, will be used to clarify the scoring scale during scorer training.
- After an **intensive training** session as described earlier in this proposal, qualifying rounds will be conducted by Scoring Directors.
- **Qualifying** responses are similar to training examples in that they have been pre-scored by rangefinding committee members. The responses will be divided into sets and scored independently by each scorer trainee. The data from these qualifying rounds will be used to determine which scorer trainees will be qualified for actual scoring.
- **Recalibration** responses may be used throughout the scoring session to monitor the scoring by comparing each scorer’s scores to pre-determined scores. Similar to the training and qualifying materials, the recalibration materials will be selected from responses scored by the rangefinding committee. Recalibration sets will be used to re-focus scorers on the

Pennsylvania scoring standards by comparing the pre-determined score to that assigned by the scorer. In addition, these examples may be used by the Scoring Director or Team Leaders for a retraining session.

- **Validity** responses detect possible room drift and individual scorer problems. Validity reports compare scorers' scores to pre-determined scores. The validity responses are "blind" to the scorers: scorers cannot distinguish a validity response from any other type of response.
- Team Leaders will conduct routine **read-behinds** for all scorers.
- Another measure of rating scoring quality is **inter-rater reliability and score point distribution reports**. To monitor scorer reliability and maintain an acceptable level of scoring accuracy, DRC will closely review reports that will be produced daily. The reports document individual scorer data, including scorer number, number of responses scored, individual score point distributions, and exact agreement rates. DRC will investigate the issue and resolve any problems those reports identify. DRC will provide these reports to PDE on a regular basis.

DRC is proud of the scorer quality-control reports that we have developed for our handscoring projects. Because we can produce these reports on demand, we can assure PDE that immediate action will be taken to resolve scoring discrepancies within minutes (when necessary) of the first and/or second reading. A complete description of our quality-control reports can be found under *Subheading VII.I.13*.

Monitoring

During the handscoring process, the Scoring Directors will meet with their Team Leaders each morning to review the statistics generated from the previous day's work. If scoring patterns are apparent among scorers, each Team Leader will deal with these issues on an individual basis. Our imaging system allows a Team Leader to determine read-behind rates (frequency of monitoring) for each scorer. If the scorer needs clarification of scoring rules, or is scoring tentatively, we typically monitor one out of five readings. The imaging system randomly selects which images the Team Leader will read behind.

We also will study the inter-scorer agreement. If a scorer falls below 70% perfect agreement or assigns nonadjacent score points for English Composition responses, the Team Leader will re-train the scorer and look for improvement within 24 hours. If the inter-scorer reliability is considerably lower than the agreed-upon rate, DRC will remove all assigned scores given by the scorer in question. The images will then be re-dealt and rescored.

If the scorer shows no improvement in the 24-hour period, we will remove the scorer from the project. All scores from the terminated scorers will be erased, and the responses will be re-routed through our Imaging System.

We do not report on scorer performance after the fact, as some contractors do. We believe that scorers with less-than-acceptable scoring patterns must be identified

immediately and those patterns corrected. DRC has worked diligently to devise effective monitoring reports and procedures to accomplish both detection and correction. If a scorer is terminated or is given the opportunity to re-train, our imaging system erases all scores from that scorer and re-deals those images to qualified scorers.

Scoring Difficult Responses

Our imaging system enables a scorer to forward difficult responses to the Team Leader. Also, since our teams are relatively small, quite often a scorer will raise his/her hand with a scoring question and the Team Leader will simply read the response at the scorer's imaging station. Similarly, if a Team Leader encounters an unusual response and seeks advice from the Scoring Director, he/she can forward these responses electronically.

If a scorer has trouble scoring a response because the image is written too lightly, the scorer can request a copy of the student response from the original. These responses are automatically routed to Team Leaders who verify that the original responses need to be made available. When a Scoring Center requests an original, it is sent overnight to the Scoring Center and then returned to the warehouse as soon as the scores are entered into the imaging system.

Handling Alert Papers

Unusual or aberrant responses that cannot be assigned a score will receive a nonscorable code. PDE and the rangefinding committees will define what constitutes a nonscorable response. During scoring, DRC will contact the designated PDE representative to obtain a ruling on responses that cannot be assigned a score based on our current understanding.

To handle possible alert papers (student responses indicating potential issues related to the student's safety and well-being that may require attention at the state or local level), DRC's imaging system gives scorers the ability to alert questionable student responses. Alerted images are routed to the Scoring Director who will print the response. Next, these alerts are reviewed by the Project Director, who then sends copies of the student's responses to the appropriate school district official. PDE is notified that this information has been sent, but does not receive the student's responses or any other identifying information. At no time during scoring do scorers have access to demographic information on any students participating in the assessment.

Currently, our alert system enables readers to flag papers for suspected teacher interference or plagiarism. These papers go through the same process as the other alerted papers. A Teacher Interference/Suspected Plagiarism Report may be provided to PDE at the completion of scoring. Since this was not requested in the RFP, DRC has not included costs for this service in this proposal. However, the inclusion of this service could be negotiated upon contract award.

Scoring Field Test Open-Ended Items

DRC will be responsible for all activities associated with scoring all stand alone and embedded field test open-ended items. After each rangefinding session, the Scoring Director will select responses to be used for training and will prepare necessary materials (scoring guides with annotated anchor papers, training sets, and qualifying sets).

Prior to scoring, all answer documents will be separated by form. DRC will score by unique item only. In other words, scorers will train and qualify on one field test item at a time. This is how we currently score the PSSA field tests and it is a very effective way to assess field test responses. Upon completion of scoring each field test item, DRC will summarize the different types of responses that are elicited by each item and will assist in revising the items, if necessary. Our scoring procedures are standardized and are the same for field tests and operational assessments. Additionally, DRC will use the same staff members who have been central to scoring the PSSA in the past.

VII.I.11. English Composition Scoring

As stated in the previous subheading, all English compositions will be scored twice by two independent readers. If a score is non-adjacent, a third, expert reader will score the response to resolution.

VII.I.12. Demonstration of Scanning and Reporting Procedures

DRC welcomes the opportunity to demonstrate our comprehensive and accurate scanning and reporting procedures. We understand that up to two PDE staff members may attend this demonstration, which will be performed no later than four weeks prior to each assessment administration. DRC will assist PDE in arranging onsite reviews of DRC's operations activities related to the scanning and reporting process. We will submit the results of the demonstration to PDE for process review and approval.

VII.I.13. Summary Reports from Open-Ended Scoring Sessions

DRC will prepare a number of reports to monitor the quality and effectiveness of various aspects of the project (sample reports are shown in *Appendix 6*). We will work with PDE to determine which reports are desired for the GCA Program. The reports are described in the following table.

Quality-Control Reports

Report	Report Specifics
<p>Scorer Monitor Report</p>	<p>Monitors how often scorers are in exact agreement and ensures that an acceptable agreement rate is maintained. This report provides daily and cumulative exact and adjacent inter-scorer agreement and the percentage of responses requiring resolution (if required). The calculations for this report are as follows:</p> <ul style="list-style-type: none"> ■ Percent Exact—total number of responses by scorer where scores are equal, divided by the number of responses that were scored twice. ■ Percent Adjacent—total number of responses by scorer where scores are one point apart, divided by the number of responses that were scored twice. ■ Percent Non-Adjacent—total number of responses by scorer where scores are more than one score point apart, divided by the number of responses that were scored twice.
<p>Score Point Distribution Report</p>	<p>Monitors the percentage of responses given each of the score points. For example, for the open-ended responses, this daily and cumulative report shows how many 0s, 1s, 2s, 3s, and 4s a scorer has given to all the responses he or she has scored prior to the time the report is produced. It also indicates the number of responses read by each scorer, so that production rates can be monitored.</p>
<p>Item Status Report</p>	<p>Monitors the progress of handscoring. This report tracks each response and indicates the status (e.g., “needs a second reading,” “complete”). This report ensures that all discrepancies are resolved by the end of the project.</p>
<p>Responses Read by Scorer Report</p>	<p>Identifies all responses scored by an individual scorer. This report is useful if any responses need rescoring due to potential scorer drift.</p>
<p>Read-Behind Log</p>	<p>Used by the Team Leader/Scoring Director to monitor intra-rater reliability. Team Leaders read a random selection of scored responses from each team member. If the Team Leader disagrees with the scorer’s score, remediation occurs. This has proven to be a very effective type of feedback because it is done with items live-scored by a particular scorer.</p>
<p>Recalibration/ Validity Reports</p>	<p>Provided daily. These two processes are conducted throughout the entire scoring process. Both processes compare pre-determined scored responses to scorers’ scores for the same set of responses. Additional responses are given to individuals if the Scoring Director feels that it is warranted.</p>

VII.I.14. Annual Scorer Drift Studies

Open-ended (OE) items are a component of many large-scale assessments, as they will be for the GCAs. When scored by trained raters, using well-established protocols, OE items have strong inter-rater agreement within project, rarely below 65% exact agreement and 90% exact and adjacent. Values above 98% exact and adjacent are not uncommon. The number of score points and content area are the main factors that affect agreement rates. (Mathematics OE items typically show higher exact agreement rates than Reading OE items.)

Inter-Rater Agreement for 2006 Grade 5 PSSA Constructed Response Items

Item	% Exact Agreement	% Adjacent Agreement	% Exact + Adjacent Agreement
Reading			
1	74	25	99
2	70	29	99
3	70	29	99
4	71	27	98
Mathematics			
1	79	21	100
2	96	4	100
3	90	10	100

Even with such high levels of agreement, the assigned scores can still be influenced by random and systematic sources of error. Some potential sources include how the rubrics assign points, what the raters are trained to check, and which papers are chosen as exemplars to define each score point. There may also be extraneous environmental or societal influences that affect the score that might be assigned to any particular work sample in any given year.

Some level of a between-year effect is expected from this *construct-irrelevant* variance, and it will make linking across years more problematic and score interpretation less general. However, it need not be fatal to the assessment.

The following section is a description of DRC’s approach to detecting and managing reader drift. First, it is important to clearly define *rater drift* and *rater effects*. **Rater drift** refers to how one group of raters changes their severity within the same scoring window (rater differences within year). *Does an individual rater become more or less severe in applying the rubric as more papers are scored?* This is internal consistency within a year. **Rater effect** refers to differences in the severity of how two different groups of raters score the items (rater differences across years). *Will the same paper receive the same score from this year’s raters as it did from last year’s raters?* This is consistency across years.

Preventing Reader Drift

DRC utilizes a handscoring **validity process** to prevent *reader drift*. See the previous subheadings for more on DRC’s handscoring quality process. The goal of the validity process is to ensure that scorers are trained to a specified standard, a standard that is applied across years. This is accomplished by having readers score student responses that have “*criteria*” scores; scores that have been previously assigned by two previous scorers and confirmed by DRC’s scoring content specialists.

The validity process begins with the selection of scored responses. As an initial pool, DRC selects approximately 200 previously scored responses for each item that include representation of all score points from the item’s rubric. The content specialists for each subject then go through the selected responses and pare it

down to approximately 40 examples. Papers that represent different types of responses within the same score point are used to ensure that readers understand the full spectrum of each score point. These selected responses are distributed periodically to the readers, who are not aware that they are scoring criterion papers.

The readers' assigned scores on these validation papers are compared to the previously determined (criterion) scores to determine the "validity" of the scores from these new readers. For each item, the percent of perfect agreement and the percents of high and low scores are computed. This is called the *Validity Item Detail Report*. The same statistics are also computed for each reader. This report is the *Validity Reader Detail Report*. The reports are produced twice a day. This includes both daily and cumulative reports.

If the rate of agreement for a reader on the validity paper scores falls below an acceptable level, the scores are purged for that reader from that day and the papers re-circulate to be re-scored by other readers. An acceptable rate of agreement is typically defined as below 70% perfect agreement. The reader in question is either retrained or dismissed.

DRC is constantly looking for patterns in scoring that may need correction and the cumulative report is quite useful for this purpose. For example, there may be a validity response that is a score of three, but has received the score of two by the majority of scorers. The DRC content specialist will then find similar types of responses which reinforce the explanation of the three score point for that particular type of response. Room-wide training will then occur. Often, similar types of responses can be found in the training material and a quick review of specific training papers is enough to get the readers re-aligned.

The validity process is an additional training/quality process conducted by DRC's Performance Assessment Services (PAS). This quality assurance process is conducted after the readers go through rigorous training and have demonstrated that they have qualified to be readers. Additionally, PAS conducts calibration training (i.e., readers as a group are retrained to score responses with known scores), on an as-needed-basis, to keep readers from drifting and to maintain integrity for the scoring project.

For the GCAs, DRC will conduct the validity process for all assessments. All available **common OE** items will be used in the validity process. For each validity item, **40 validity papers** will be selected from across the range of possible scores. The validity papers will be selected from a previous year's administration. In short, any reader that qualifies to score for a particular course will be given 40 validity papers to score, distributed throughout the scoring window.

The 40 validity papers for each item will be selected to roughly **match the score distribution** of the item from the prior year, with a constraint of a **minimum of 5** papers per score point.

Validity responses are “blindly” routed to the readers so they should not be aware that they are scoring validity papers. This method prevents readers from being affected by knowing that their scores are being compared to pre-assigned scores, as with the scoring of qualifying responses.

The results of the validity papers will be monitored **twice daily** according to DRC’s standard process. If DRC sees that the agreement rate for any item is low, all readers will be retrained as a group to prevent further drift. Additional training may be given at the discretion of the Scoring Directors.

Preventing Reader Effects

DRC utilizes a comprehensive process to monitor and control *reader effects*. This includes standardized and consistent training of readers across years, the use of calibration sets to retrain readers, and the implementation of the validity process. In spite of these quality assurance processes, the presence of reader effects (i.e., score instability) may still occur across years. The following examples are a few situations where instability could be inherent in the process, although the items may be valid stimuli in all cases:

- An open-ended question that is field-tested with a large number of others may not have the same richness of responses available for choosing the training materials as are available after a statewide operational assessment. A field test may produce fewer than 1000 responses to each item while the operational assessments typically produce tens of thousands. In particular, there often are no more than a handful of papers at the extreme score categories for a field test. The limited number of choices could make the rater training less than optimal.
- Items that are more or less topical when originally tested may later yield responses of a different quality. For example, a social studies question dealing with the mechanics of the Electoral College would probably have been more difficult before the 2000 Presidential election than it would be immediately after it.

To determine whether readers are stable across years, one can compare the scores assigned to the same set of papers across years. For the GCA exams, the common OE items, linking the current and previous assessments, will be used for this purpose. A subset of those student responses (with scores assigned by the previous year’s raters) will be distributed blindly to the current year’s raters.

DRC has conducted similar rater effects studies on the PSSA for PDE. In those studies, a power analysis established that approximately 600 papers are needed to detect the presence of a rater effect if a power of .80 was desired, assuming an effect size of .10 and an alpha level of .10. For the GCA rater effect studies, 600 papers are proposed per exam.

The goal is to implement the scoring of the previous year’s responses with minimal impact to the scoring of the current year’s responses. For every n-th

paper scored, the readers will receive one rater effect study paper (i.e., response from the previous administration) to score. The exact rate will be based in part on the number of readers available, the length of the scoring window, and the capabilities of the scoring system.

For the rater effect studies, the following statistics will be calculated for each item:

- Mean and standard deviations of the item score for each year.
- t-test of the mean differences, showing effect size (ES), t-statistic (T), and *p*-value (*p*).
- Correlation between the scores from previous and current assessments.
- Agreement rates between the previous and current scores, showing:
 - Percent exact agreement,
 - Percent exact + adjacent (Exact+Adj),
 - Cohen's Kappa (Kappa),
 - Weighted Kappa (Weighted K), and
 - Brennan and Prediger's Kappa (B & P) rates.

Accounting for Rater Effects in the Calibration and Linking Process

For the GCAs, DRC proposes to use a similar approach that is currently used for rater effects in the equating process of the PSSA. This approach has been presented to PDE's National TAC members.

When the difference between the means of the previous and current paper scores is **statistically significant**, the reader effect will be estimated and the adjustment applied to the thresholds for the OE items. The effect will be estimated with a concurrent calibration of the item with both the current and previous scores. Because it is the same item in both cases, any difference in the calibrations will be ascribed to the rater effect. The adjusted OE item parameters values will be used in the current year's equating process (e.g., in calculating the mean shift equating constant).

The above process was proposed for the post-equating process conducted on the PSSA. This process can also be implemented with the plan to pre-equate the GCAs. In this context, it will be implemented during the calibration and linking of the field-test items to place them on the GCA item scale.

For the GCAs, DRC is committed to conducting both the validity process and the rater effect studies to investigate consistency of scores within a year and consistency of scores across years.

VII.I.15. Report of the Open-Ended Scoring Process in the Annual Report

Based on the records kept and reports generated during the handscoring of the GCA Program, DRC will include a documented report of the open-ended scoring process in the annual technical report each year of the contract.

VII.I.16. Storage, Retrieval, and Destruction of Materials

Upon completion of processing, secure materials, including used test booklets and answer documents, will be securely boxed and stored at DRC's secure processing facility for a period of one year after schools and districts receive their test results.

Using DRC's proprietary Operations Materials Management System (Ops MMS), we can quickly and efficiently retrieve processed and stored test booklets and answer documents as the need arises, either during or upon completion of processing. Individual student tests will be retrieved within five business days of DRC's receipt of official PDE requests. The following steps will ensure the quick retrieval of documents:

- Project-specific box labels will be created containing the following information, as applicable: unique customer and project information, materials type, batch number, pallet/box number, and the number of boxes for a given batch.
- Boxes will be stacked on project-specific pallets. Each pallet will be labeled with a list of all the batches it contains.
- Before each pallet is stored, a quality check will be performed to ensure accurate boxing and pallet content labeling.

Electronic images and data will be stored for the life of the contract plus one year beyond. The storage system will allow efficient and easy retrieval of individual student tests within a short timeframe.

All hard-copy materials and electronic images and data will remain secure until written authorization has been received from the appropriate PDE contact to release or securely destroy the documents and files.

Materials Retrieval

All retrieval requests will be submitted to PDE for approval prior to processing. After approval is received, DRC staff will initiate and track the retrieval process. Processed test booklets and answer documents can be retrieved quickly and efficiently as the need arises, either during or upon completion of processing. Individual student tests (original hardcopies) will be easily retrievable because of DRC's effective document storage procedures (please see above). Additionally, DRC's IBML image scanners and Image Scoring System allow for on-demand retrieval of specified images (e.g., specific batch files, specific courses, specific

students); each image is assigned a unique identification number that allows for quick and easy retrieval at the student and school level.

Depending on PDE preference, either paper or electronic copies of individual student tests can be provided to the requesting party. Electronic copies would be available as PDF files on CDs or other desired media; the PDFs would be viewable using Adobe Acrobat Reader.

DRC's process for the retrieval of individual student answer documents is as follows:

- Project Management supplies the Materials Processing Manager with the detailed information, including the security number of the document to be retrieved.
- The Materials Processing Manager enters the security number into DRC's secure inventory control system to identify the building, rack, pallet, and box location of the document.
- The Materials Processing Manager retrieves the document from storage and forwards it to Project Management.
- A placeholder document is inserted into the box identifying the material that was removed and the booklet is assigned via security number to Project Management in the inventory control system.
- A copy of the request is maintained on file in the warehouse for future reference.
- Project Management forwards the requested document to PDE and/or other parties as directed by PDE.

Requests for Reprocessing

All rescore or reprocessing requests will be submitted to PDE for approval prior to rescore processing. After approval is received, DRC staff will initiate and track the retrieval and rescore process. All rescoring will be scored manually, including multiple choice items, by experienced and qualified personnel. All applicable security and quality-control procedures that were implemented by DRC during the original processing and scoring will be maintained. Rescoring will be completed within 45 business days of receipt of PDE rescore approval. Reprocessing and rescoring will be available for 120 days after schools and districts receive their test results. DRC reserves the right to charge for rescore requests, except in the event that any materials have been inaccurately processed, in which case DRC will retrieve and reprocess them at our own cost.

VII.J. ANALYSIS OF ASSESSMENT DATA

VII.J.1. Ensuring Comparability of Scores

DRC has many procedures in place to ensure comparability of scores across years. Using a combination of calibrating, scaling, and equating, DRC will build for the Commonwealth, a strong link to the previous test administrations ensuring comparability across years.

VII.J.2. Calibration and Scaling Procedures

Rasch Measurement

In order to derive data worthy of the name measurement, Georg Rasch reasoned that one person parameter (ability) and one item parameter (difficulty) must govern the interaction between the person and the item. If the person has more ability than the item has difficulty, the person would have a higher chance of answering correctly. If the person has less ability than the item has difficulty, the person would have a higher chance of answering incorrectly.

This line of reasoning led to the simple logistic model, which has several closely related and very useful properties:

- Mathematical separability of the model parameter.
- Sufficient statistics that do not involve the parameters.
- Specific objectivity in the measurement, sometimes called person-free calibration and item-free measurement (Wright, 1968).

Specific objectivity means that the estimation equations for ability do not involve the difficulty parameters, and the equations for difficulty do not involve the ability parameters. In practical terms, this means that students can be ordered along the measurement continuum by their number correct scores and that items can be ordered along the continuum by the number of correct answers to the item. No other information is necessary and anything remaining in the data can be used for control of the model. *Specific objectivity* is the cornerstone of the Rasch family of measurement models.

Because both right-wrong and open-ended (OE) items are part of the assessments, DRC proposes to use a common item calibration that places both item types onto a common scale. The multiple-choice items (MC), scored right or wrong, will be calibrated using the familiar form of the Rasch model (Rasch, 1960). The open-ended items will be calibrated using the partial-credit Rasch model (Wright & Masters, 1982; Masters, 1982). The latter model parameterizes the thresholds needed to reach a complete solution to the task so that there is one item difficulty parameter associated with reaching each possible score after zero. A multiple-choice item may be thought of as a partial-credit task with only a single threshold.

The Rasch model applicable to dichotomously scored items (MC) can be expressed in the most familiar form of the model:

$$1. \quad \Pr(\text{right} \mid \beta_n, \delta_i) = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}.$$

The probability of success for a person with ability β_n on an item with difficulty δ_i is determined by the difference between the ability of the student and the difficulty of the item.

With the partial-credit model used for open-ended items, π_{nik} is the probability that person n will score k on item i . Then, the *first* threshold for item i is a score of 1 rather than a 0, which is the conditional probability of a score of 1, given a score of 0 or 1:

$$2. \quad \Phi_{1ni} = \frac{\pi_{ni1}}{\pi_{ni0} + \pi_{ni1}} = \frac{\exp(\beta_n - \delta_{i1})}{1 + \exp(\beta_n - \delta_{i1})},$$

where β_n is the ability of person n and δ_{i1} is the difficulty of the first threshold. The expression on the right is identical to the Rasch model for a dichotomous item. The only differences are that now $\pi_{ni0} + \pi_{ni1} < 1$, since more than two response categories are provided, and δ_{i1} , while still the difficulty of the first threshold for item i , is not the difficulty of the only threshold for the item.

For example, with a three-point open-ended item, where a person n must make one of the four possible scores (0, 1, 2, or 3) on item i ,

$$3. \quad \pi_{ni0} + \pi_{ni1} + \pi_{ni2} + \pi_{ni3} = 1$$

These relationships can be rearranged to obtain one general expression for the probability of person n scoring x on item i :

$$4. \quad \pi_{nix} = \frac{\exp \sum_{j=1}^x (\beta_n - \delta_{ij})}{1 + \sum_{k=1}^{m_i} \exp \sum_{j=1}^k (\beta_n - \delta_{ij})}, \quad x = 1, \dots, m_i$$

If the number of thresholds (m_i) is one, the summations in expression (4) drop out and it reduces to expression (1).

Joint-maximum-likelihood estimation of items will be accomplished using WINSTEPS (2008). This calibration software is commercially available and widely used in the testing industry. The capabilities of the WINSTEPS program will be utilized to assess unidimensionality, item interdependence, and other deviations from the model. The program has many options for the exploration of the person-item residual matrix (Mead, 1976; Smith, 2000; Ludlow 1986).

An important consequence of these models is that the number of correct responses to a given set of items is a sufficient statistic for estimating person ability. As a result, each person with the same raw score will be assigned the same estimated ability.

Rasch calibration and scaling has become a relatively routine operation. To accomplish the necessary steps, DRC will employ WINSTEPS, which provides estimates of the item and threshold difficulties, as well as a plethora of fit and diagnostic statistics.

VII.J.3. Equating Procedures

Rasch Equating

No equating methodology can compensate for a poor test design. Angoff (1971) outlined four conditions that must be satisfied for equating to succeed:

- The test forms to be equated should measure the same ability (unidimensionality).
- The resulting raw score to scaled score conversion should be independent of the data used in deriving it and should be applicable in all similar situations.
- Scores on the two test forms should, after equating, be interchangeable in use.
- The equating should be symmetric, or the equivalent, regardless of which test form is designated as the base.

Conforming to the Rasch model, which is the basis of the GCA, ensures that these conditions are satisfied.

Both content and difficulty matching are prudent components within an overall test design that allows for a given scaling and equating design to succeed. To succeed, in this sense, means that after equating, a student's score may be compared to another's within or across years in an equitable and objective manner.

Test forms that conform to Rasch's principles are assured of satisfying Angoff's requirements. That being said, building forms to this standard is a challenging and on-going task. It requires the careful development of items, as described in the item and forms development sections, to ensure the content of the items is consistent with the content standards, the curriculum, and the instruction. It also requires strong statistical controls to ensure that all items are equally valid and reliable instances of the underlying construct. Because of the strict requirements of the Rasch model, it is the ideal vehicle for providing these controls. Strict adherence to the model's requirements will be the guiding principle to develop sound measurement scales, to maintain consistent performance standards, and to facilitate comparable reporting, across forms and across years.

The specific objectivity property of the model allows, once the sufficient statistics have been removed from the scores, the remaining data to be used for control and monitoring. The data should have no lingering influences dependent on the distribution of ability in the group who provide the calibrating sample, nor on which administration is considered. Any patterns related to years or groups will be dissected to determine how the differences arise. Like most forms of data analysis, the statistics will be used to call attention to problematic situations, but the substantive interpretation will require the collaborative efforts of educational specialists, PDE, and the TAC.

Much of these analyses will draw on the processes described in the statistical and psychometric analyses discussed in *Subheading VII.B.9*, as well as what is discussed below. It will also take advantage of methods from both traditional true score theory (Angoff, 1971; Kolen and Brennan, 2004) and IRT (Cook & Eignor, 1991).

Implications of Open-Ended Tasks on Equating Procedures

Theoretically, the Rasch equating procedures should be independent of item type and the same general process of computing mean logit difficulties and displacements are applied throughout. However, there is some evidence, seen in differences in item-trait correlations and within group standard deviations, that the constructs tapped by the different item types are not identical. This is an argument for why open-ended tasks (OE) should be both included on the assessment and represented in the link. It is also an argument for why the performance of the OE items must be carefully monitored in the scoring and equating process.

The first, practical, question is how to include the OE logit difficulties. They can be introduced either as a single location parameter for each item, as with a rating scale that uses an item location and off-sets from there to describe the thresholds, or as a unique difficulty for each threshold for a task, as with the partial credit model. For purposes of describing the responses, there is no difference between the two views, but they are quite different for purposes of equating.

It has been DRC's experience that the former view, which uses a single location for each task, results in a more stable equating process and is more consistent vis-à-vis multiple-choice items (MC). The estimates of the extreme threshold parameters are often based on very small numbers, have much higher standard errors, and frequently dominate the entire equating process. In contrast, the mean of the thresholds for a task (i.e., the single location estimate) is much more stable and should contain the relevant information from the OE, but with the error better managed.

For these reasons, DRC considers the OE items an important component of the equating process and proposes to include them by using the task location estimates as additional link items. They will then be subject to all the same outlier and displacement analyses as the MC.

Importance of Consistent Item Sequence

In maintaining consistent metrics and performance standards across administrations and years, it is essential that the overlapping items be presented in a consistent context. This is especially true in preserving the metric. The effective difficulty can be affected by location on the test. There may be start-up effects that cause students to under-perform on early items, resulting in over-estimates of the difficulties of these items. There may also be fatigue factors that have a similar effect on items late in the test. During test development, DRC will make every effort to ensure the position of items on the operational forms will be similar to their location on the field test.

This effort will be greatly aided by the use of fixed embedded field test positions. In addition to improving the integrity of the field test by making the testing context identical to the operational and by making it unlikely that the items will be identified as field test, it also makes it possible to try the items out in locations similar to their eventual placement on the operational.

In summary, DRC is aware of the issues surrounding the importance of the testing context in general and will strive to ensure these factors do not unduly influence results of the assessment.

Equating

To meet the accelerated reporting schedule requested by PDE, DRC is proposing an efficient equating design that allows for the quick turnaround stated in the RFP. Under this design, existing item parameter estimates are used to compute the raw-to-scale conversion table for a test as soon as it is constructed. There is no need to wait until the test administration to create the scoring tables. DRC has successfully used this equating design in other states, so we are confident that this approach will work for the GCA Program. **DRC proposes to use this equating procedure for all test administrations (i.e., spring, summer, fall).**

DRC's equating procedure is based on the expected number correct derived from the Rasch scaling model. Here, the association between raw scores and abilities defines the raw to scale score conversions needed for scoring and reporting. For a given ability, the expected raw score (r) on a test is simply:

$$1. \quad E(r) = \sum p_{ij}$$

where p_{ij} represents an expected item score. For multiple-choice items worth one point, this is equivalent to the probability that a student will answer the item correctly, which is given by:

$$2. \quad p(x_{vi} = 1 | \beta_v, \delta_i) = \frac{e^{\beta_v}}{e^{\delta_i} + e^{\beta_v}} = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}}.$$

For polytomous items, the probability of receiving an item score of k is given by:

$$3. \quad p\{k | \beta_v, \delta_i, (\tau_{j=1,m})\} = \frac{e^{\sum_{j=1}^k (\beta_v - \delta_i - \tau_j)}}{1 + \sum_{x=1}^m e^{\sum_{j=1}^x (\beta_v - \delta_i - \tau_j)}} = \frac{e^{k(\beta_v - \delta_i) - \sum_{j=1}^k \tau_j}}{1 + \sum_{x=1}^m e^{x(\beta_v - \delta_i) - \sum_{j=1}^x \tau_j}}$$

In short, the expected total test score is equal to the sum of the expected item scores over all operational items.

Equating Procedure

DRC will use WINSTEPS to generate the conversion tables. DRC psychometricians are very familiar with WINSTEPS output formats and have experience handling multiple output files programmatically. Psychometric staff will run WINSTEPS by anchoring all item difficulties and thresholds. WINSTEPS raw-to-measure conversions are completely model driven in a ‘fully anchored’ run. Raw-to-measure tables will be obtained from the subsequent WINSTEPS output files. These files will then be checked using independent procedures (see *Subheading VII.I.3* on psychometric quality procedures). A sample WINSTEPS scoring table is presented below. The raw score to measure conversions are provided in the first two columns. The measure scores are linearly transformed to derive the GCA scale score.

TABLE 20.2 Reading G8 GACRCT **Anchored** CAL. GA08REA.OUT Jun 23 10:41 2003
INPUT: 39559 PERSONS, 40 ITEMS MEASURED: 39559 PERSONS, 40 ITEMS, 2 CATS 3.37

TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST

SCORE	MEASURE	S.E.	NORMED	S.E.	FREQUENCY	%	CUM.FREQ.	%	PERCENTILE
0	-5.2840E	1.8367	-28	151	0	.0	0	.0	0
1	-4.0520	1.0199	74	84	0	.0	0	.0	0
2	-3.3186	.7350	134	61	1	.0	1	.0	1
3	-2.8728	.6114	171	50	0	.0	1	.0	1
4	-2.5445	.5393	198	44	6	.0	7	.0	1
5	-2.2803	.4911	220	40	19	.0	26	.1	1
6	-2.0566	.4565	238	38	50	.1	76	.2	1
7	-1.8604	.4303	255	35	77	.2	153	.4	1
8	-1.6841	.4099	269	34	109	.3	262	.7	1
9	-1.5229	.3936	282	32	194	.5	456	1.2	1
10	-1.3733	.3804	295	31	226	.6	682	1.7	1
11	-1.2328	.3696	306	30	296	.7	978	2.5	2
12	-1.0996	.3607	317	30	340	.9	1318	3.3	3
13	-.9721	.3534	328	29	376	1.0	1694	4.3	4
14	-.8494	.3475	338	29	496	1.3	2190	5.5	5
15	-.7303	.3427	348	28	491	1.2	2681	6.8	6
16	-.6142	.3389	357	28	558	1.4	3239	8.2	7
17	-.5004	.3361	367	28	671	1.7	3910	9.9	9
18	-.3881	.3341	376	28	673	1.7	4583	11.6	11
19	-.2770	.3329	385	27	720	1.8	5303	13.4	12
20	-.1663	.3325	394	27	829	2.1	6132	15.5	14

Item Bank Maintenance

Monitoring and updating item calibration values to control for issues such as item parameter drift can help establish and maintain a successful equating program.

DRC recommends the following procedures be used to ensure that the most appropriate item difficulty parameters are “banked” for later use:

- Using the full data file, conduct a free/local calibration for all operational items.
- Evaluate the stability of the local calibration results vs. the “banked” difficulties using Robust Z analyses.
- Using only values for “stable” items, determine the *mean shift*.
- For operational items, update bank values by applying the mean shift to all operational items so as to put the items on bank scale. A viable alternative would be to weight the item difficulties by their respective item information. In either case, the resulting transformed Rasch difficulties will be banked and applied in future applications.

Calibrating Embedded Field Test Items

To calibrate embedded field test items, DRC recommends the use of a fixed common item parameter (FCIP) procedure to anchor the estimates for all operational items and then estimate the Rasch item difficulty parameters for the field test items. In this equating design, this entails anchoring all operational items to their banked values determined after the maintenance activities described above have been performed. This approach puts all field test items on the operational scale regardless of the form on which they were administered.

VII.J.4. Item Analyses

Based on meeting previously discussed psychometric guidelines (*Subheading VII.C*), items that do not go through item review are immediately placed in the item bank. Items that go to item review will be placed in the bank if approved or made available for retest if revised. Once analysis on the field test items is completed, the items are eligible for use on operational forms. Please see *Subheading VII.C* for information on forms construction.

Construction of the GCA Module Exams

To begin construction of the GCA module exams, items will be field tested in the Spring of 2010. This standalone field-test event will be designed to generate sufficient items to create the first operational module form for use in Fall (Dec.) 2010 and the parallel forms for the following spring, summer and fall administrations (for Algebra I, Algebra II, and Geometry). Sufficient items will be included to allow for the creation of the necessary breach forms. The items from the Spring 2010 field-test event will be subjected to all the same analyses that have been discussed in *Subheading VII.B.9*.

The figure below illustrates DRC’s plan for form construction. It shows the field-testing event in Spring 2010 and how the forms for the following years are constructed. Future parallel operational module forms will be developed using items from the calibrated item bank. Most of the items will come from the

previous spring administration’s embedded field-test pool (noted by the solid arrows in the figure), save for some operational items selected to ensure test specifications and target test information functions are matched (noted by the dashed arrows in the figure). Recall from a previous section that the fall and summer module forms will include field-test items. However, those field-test items will not be used to construct new forms; they are included only to keep the student testing experience similar across the three test administrations.

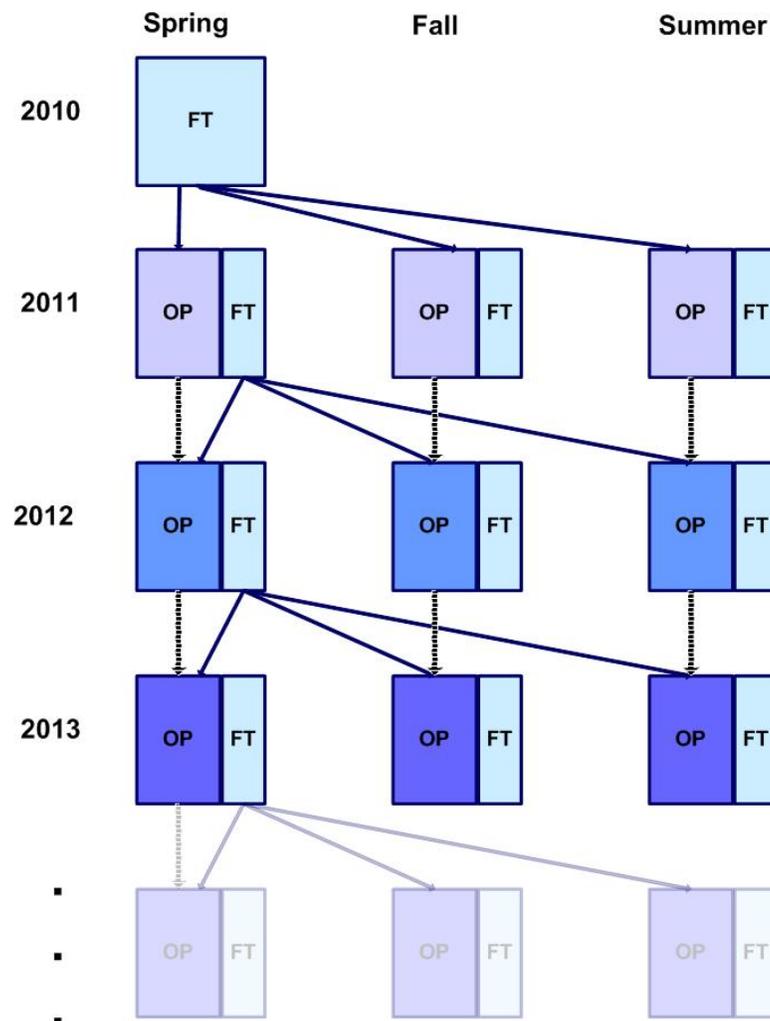


Illustration showing how the field-test (FT) items are tried out and used to create the operational (OP) forms

Full population data from the operational Spring 2011 administration will be calibrated to set the item bank scale. These data will provide the basis for the standard-setting process, which will lead to establishing the performance-level standard and defining the final reporting metric (i.e., the scaled score). Future embedded field-test items will be scaled to be on the bank scale, such that future forms for each module can be developed using items from the item bank.

Operational Items

Item analysis of operational items is shown in the chart below. Analyses will include the following:

Classic Item Analyses (Overall and by Subgroup where Requested)	
p -values, with flags for very easy and very difficult items	Percent choosing each multiple-choice option, with flags for distractor percent higher than correct-answer percent
Corrected item-total correlations, with flags for possible mis-key or poor item quality (point-biserials)	Option-total correlations for MC items
Percent of students earning each open-ended item score	Standard error of measurement for the scale
Test reliability	
Rasch Statistics	
Item Statistics:	
Logit difficulty estimates	Step parameter estimates for OE items
Standard errors for all parameter estimates	INFIT and OUTFIT statistics
Test Indices:	
Test information function	Test characteristic curves
Raw-Logit-scale score conversion tables	Standard errors for all parameter estimates and scale scores
Person separation reliability	

Classic Item Analyses

As discussed previously in *Subheading VII.C*, DRC will use its proprietary Item and Test Evaluation Modules (*iITEMs*) for classical item analyses of operational items. The key verification module of this system computes the number and proportion of students selecting each response option, the p -value for the item, the item-total correlation (e.g., the point-biserial correlation) for the key, and the item-total correlations for each of the other response alternatives. These statistics are used to flag any potentially incorrect scoring keys. DRC psychometricians will work with PDE to define the criteria that are most suited for the GCA Program. Also discussed in the previous section, the exact criteria for flagging an item can be customized within *iITEMs*.

In the *distractor* analysis module, *iITEMs* generates a graph depicting the proportion of students selecting each response option as a function of raw score. The proportion of students selecting the keyed response option should increase as a function of ability. Conversely, the proportion of students selecting each of the distractors should decrease as a function of ability. A graph for an item that does not show this pattern of results may indicate an incorrect key. DRC has found that these item distractor analysis graphs, in conjunction with the traditional item statistics, are powerful tools in detecting possible item mis-keys.

The item analysis will be conducted as soon as data based on a large calibration sample is available. This analysis will be conducted by form. All items flagged as

possible mis-keys will be referred to DRC content specialists, Project Management, Information Systems, and Software Quality Assurance staff for further review and verification. Possible incorrect item keys will be identified, confirmed, and corrected before the final scoring is conducted. Therefore, there will be no implications for item calibrations, scaling, equating, and reporting. Documentation related to any item discrepancies and a copy of the item analysis will be available to PDE for review upon request.

Rasch Statistics

Rasch statistics will be calculated using *WINSTEPS*, a commercially available software package commonly used in the industry. For all items, threshold difficulty parameters will be provided with their associated standard errors of estimation. In addition, *infit* and *outfit* statistics will be provided.

VII.J.5. Assessment Construction Analyses

DRC's Psychometric Services staff play an integral role in the test construction process. A detailed discussion of the proposed test construction process may be found under *Subheading VII.C*.

VII.J.6. Analyses to Verify Accuracy of Scoring

Psychometric Quality and Methodology

DRC has a department devoted solely to ensuring psychometric quality in all of our large-scale assessment programs.

DRC's Psychometric Services (PS) Department is committed to quality and excellence. We achieve psychometric quality and excellence by assuring that our practices and procedures meet the professional measurement standards outlined in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999). **In addition, in 2006, DRC took the initiative and led the industry by starting a Department of Psychometric Quality.** One of DRC's main initiatives has been the creation of a data forensics system. This system, called **Psychometric Scenes Investigator (PSI) will systematically and efficiently conduct numerous analyses to ascertain the integrity of test results.** A more detailed discussion of PSI can be found in the proposed option for data forensics, included in *Appendix 8*.

The Psychometric Services Quality (PSQ) staff consists of research analysts skilled in research and measurement theory and methodology. Working closely with the psychometricians, they would provide psychometric quality control for the GCA through the following procedures:

Psychometrics Quality Procedures

- **Data files**—Quality checks will be performed by this team to verify the integrity of data files.
- **Scored data**—Quality checks will be performed on the data to ensure that test scores have been scored correctly.
- **Calibration, Scaling, and Equating**—Replication of these processes will be performed as an independent quality check.
- **Reports**—Validate that the assessment results are accurate and allow for valid interpretations.
- **Item Statistics**—Ensure item statistics are properly stored in the item bank system.
- **Trend Analysis**—Quality checks will be performed on the data and scoring to ensure no anomalies exist relative to historical performance trends.

Data File Quality Control

Psychometric quality begins with a check of the student response data file. All fields critical to the analysis, calibration, and equating process are checked and verified by the psychometric and the psychometric quality teams. Variables are validated against the final approved file layout and processing rules to ensure that no unanticipated values exist and that data characteristics appear to be consistent with past experience. All key demographic fields are checked for accuracy. Additionally, a reasonableness check on the data is conducted by computing the raw score frequency distribution, verifying the proper numbers of items and the proper location of open-ended items, and by verifying that no unusual values exist.

Preliminary Item Analysis for Key Verification

Psychometric quality control continues with a preliminary item analysis key check on multiple-choice items. There are many levels of key verification that take place within DRC (e.g., content experts take the exam and compare their keys with the approved scoring key and SQA staff tests and verifies the scoring program), but this preliminary item analysis serves as a final check to identify any items that do not seem to be functioning as expected. The preliminary item analysis is performed on a scored student file as soon as enough student records are available and is performed only on the multiple-choice items. Items that exceed certain criteria in terms of psychometric characteristics are flagged and re-verified by content experts to ensure that the identified item key is correct.

The preliminary item analysis is conducted by the psychometric team and verified by the psychometric quality team. The process is an integral part of ensuring quality and the validity of the test results. The analysis provides assurances that the test is of high quality, and therefore valid inferences can be made from the results.

Quality Check on Equating

For information on quality checks for equating, please see *Subheading VII.C.*

Student Data Files and Reporting Quality

Using the scoring tables determined through the equating step, the students' raw scores will be converted to scale scores based on those look-up tables. The scale scores will be applied to the final student data files, individual student reports, and summary reports. All files and reports go through multiple levels of quality checks. The PSQ research analysts serve as one part of that process by performing independent checks on the data files and reports.

Item Bank Process

As previously discussed, all calibrations will be run independently by the psychometrician and statistical analyst and subsequently verified by the research analysts. After these independent runs are conducted, the entire team will review and evaluate the results. This process will be directed by the Lead Psychometrician and the Director of Quality for Psychometric Services. Any discrepancies will be noted, discussed, and resolved before the calibrations will be considered final and imported into the item bank system.

Once statistics are put into the item bank, all data elements will be checked to ensure that they have been imported without error. Lastly, sample data cards will be printed from the system and checked to ensure that proper statistics and values are displayed correctly.

Internal Advisory Panel

DRC and its assessment partners are constantly striving to enhance psychometric knowledge and processes to create the most innovative solutions for ensuring valid, reliable, and instructionally sensitive assessments for all clients. In keeping with this goal, and as a special offering to PDE, DRC will be creating an internal Advisory Panel.

This panel will analyze enhancements or improvements to the GCA and recommend different options and solutions. It will create an environment for panel members and PDE to share, discuss, and debate the feasibility of applying new approaches and recent trends in large-scale assessment to all the Pennsylvania assessments. PDE will be able to draw on this added measure of guidance and creative thinking, as it continues to develop and administer a valid and reliable assessment system that supports instruction and accountability. DRC envisions one Internal Advisory Panel for the entire suite of Pennsylvania

assessments for which DRC is responsible (i.e., GCA, PSSA, PSSA-M), similar to the Technical Advisory Committee.

The Internal Advisory Panel will convene each year to address specific issues, including the following areas:

- Equating
- Reader drift
- Translation comparability
- Assessment reporting
- Online testing
- AYP and NCLB policy
- Special studies
- GCA
- Diagnostic Assessment Tool
- Model Curriculum

Prior to each meeting, DRC will consult with PDE on current concerns in these areas as well as any additional topics of interest to be covered by the panel. DRC also invites PDE staff to participate in the panel discussion and will create a written summary report following the meeting for distribution to all panel members and PDE. This report will identify any recommendations to be presented at subsequent Pennsylvania Technical Advisory Committee (TAC) meetings and can be shared with TAC members at PDE's request.

DRC is pleased to present the following experienced professionals as members of its internal Advisory Panel. DRC believes that the expertise of these individuals, combined with this opportunity for creative thinking and open discussion, will provide an added value to PDE.



Pennsylvania Internal Advisory Panel

Organization	Participants	Area of Expertise
DRC	Dr. N. Scott Bishop Mr. David Chayer Dr. Adisack Nhouyvanisvong Dr. Richard Smith * Dr. Ronald Mead * Ms. Pamela Hermann Dr. Melvin Webb	Psychometrics
	Ms. Patricia McDivitt	Test Development
	Ms. Alison Lyder	Handscoring
	Ms. Shaundra Sand	Project Management
WestEd	Dr. Stanley Rabinowitz *	Test Development and Psychometrics
College Board	Dr. Kathleen Williams	Test Development and Curriculum
HumRRO	Dr. Arthur Thacker*	Psychometrics
eMetric	Dr. Huixing Tang	Psychometrics and Reporting
University of Massachusetts	Dr. Stephen Sireci *	Psychometrics
CAL	Dr. John Poggio Dr. Douglas Glasnapp	Computerized Testing
HumRRO	Dr. Arthur Thacker	Psychometrics

* involved in the PSSA contract

VII.J.7. Additional Ad Hoc Analyses for Validity and Reliability Studies

DRC's Technical Report will serve as the primary vehicle for documenting reliability and validity evidence for the GCA Program. A sample Technical Report table of contents is provided under *Subheading VII.J.10*. A technical report previously developed for Pennsylvania is included in *Appendix 7*. A review of the significant reliability and validity documentation that will be provided in the GCA Technical Report is described below. Discussions at PDE's Technical Advisory Committee (TAC) meetings have frequently been the impetus for special reliability and validity studies. Examples of prior TAC studies are also provided below along with specific recommendations regarding potential future validity/reliability studies.

As part of the work scope requested in the RFP, two special studies will be conducted. These are discussed briefly at the end of this section.

Reliability

According to the *Standards for Educational and Psychological Testing (AERA, APA & NCLB, 1999)*, reliability refers to:

the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable, and repeatable for an individual test taker; the degree to which scores are free of errors of measurement for a given group.

Most achievement measures used in large-scale assessment programs report reliability coefficients that are sensitive to content sampling errors. The expected degree of score consistency across forms, that are exchangeable in content, is of great interest to achievement test users. Additionally, these coefficients of internal consistency can also be calculated from a single test administration, a characteristic that does not hurt their popularity in applied use.

Coefficient Alpha is an industry standard index and will be reported for each GCA course and module for all major population subgroups in Pennsylvania. The standard error of measurement (SEM) is also reported for all raw scores and scaled scores.

Because Coefficient Alpha is influenced by content homogeneity, a test with items targeting different domains based on the test's content specifications might reduce the value of the index. In such cases, it can be more informative to compute the stratified-alpha coefficient, a weighted reliability coefficient that accounts for the contribution of each subpart to the overall variance of the module. Stratified Alpha will result in a greater reliability coefficient than Coefficient Alpha when the covariance within domains is greater than the covariance between domains.

One might benefit from other means of stratification. For example, when a module is composed of subparts that require different cognitive tasks, like open-ended versus multiple-choice items, the traditional formula for Coefficient Alpha may underestimate test score reliability. Consideration of the assumptions and expectations for Coefficient Alpha will often suggest that it is not the most appropriate index of reliability under these circumstances. Stratified Alpha may also be employed in these circumstances to get a more appropriate indication of reliability.

For the English composition course, reliability will be reported using stratified alpha for all major population subgroups in Pennsylvania.

Classification Accuracy

Classification consistency refers to the degree with which the achievement level for each student can be replicated upon retesting using the same form or an equivalent form. Since it is not feasible to repeat GCA testing in order to estimate the proportion of students who would be reclassified in the same performance levels, a statistical model needs to be imposed on the data in order to project the consistency of classifications solely using data from the available administration. Although a number of procedures are available, DRC will continue to report two of the more well known methods, which were developed by Hanson and Brennan (1990) and Livingston and Lewis (1995) utilizing specific True Score Models.

Validity

As discussed in the *Standards for Educational and Psychological Testing*,

“...validity refers to the degree to which evidence and theory support the interpretation of test scores entailed by the proposed uses of the test” (AERA, APA, and NCME, 1999, p. 9).

There are five sources of validity **evidence** identified in the *Standards*: test content, response processes, internal structure, relations to other variables, and consequences of testing. Evidence related to content and structure is provided in most technical reports as part of standard industry practice.

Evidence Related to Test Content

For standards-based assessments, the strongest validity evidence is demonstrated in the item development and test construction processes. Developing items to the model curriculum and content standards and forms to exacting test specifications ensures that the contents of the tests are appropriate for their intended uses. These processes will be thoroughly described in the Technical Reports, as a normal part of the project documentation. The success of the process in prior years has also been verified through an external review and an “alignment study” by an independent contractor. The linking of individual items to specific instructional objectives is a powerful source of content evidence. Overviews of DRC’s item and data review meetings are provided in the Technical Reports for this purpose.

It is also crucial that students have the opportunity to learn the content. This requires that the assessments match the instruction as well as the standards and the curriculum. Upon award, DRC will work with PDE and its TAC to determine what studies should be conducted to verify the appropriate instruction is occurring.

Evidence Related to Internal Structure

In addition to the description of the item and test development process, the Technical Report will also contain evidence of construct validity. All of the following studies are a standard part of the Technical Report for the project.

Strand and inter-item correlations will be computed and interpreted with the assistance of content and curriculum specialists. As a related analysis of fit, they can support the construct validity of the assessment when they reflect the intended test structure and match relevant convergent or divergent expectations.

Descriptions of item fit to the measurement model are critical pieces of validity-related information. These will include summaries of the overall item fit statistics, which can be interpreted as evidence for (or against) construct validity. The Winsteps software provides several options for investigating the internal structure of a test. These include the overall fit statistics (Infit and Outfit), principle components analysis, and residual correlations. These will all be exploited to present a complete picture of the construct being assessed.

Analyses of differential item functioning (DIF), which is a specific form of construct irrelevant variance, will be presented as part of the standard psychometric analysis. This is again an analysis of fit and confirmation of Rasch specific objectivity. If items operate differently, comparisons across subgroups are problematic. All necessary steps will be taken to ensure the items function appropriate and that no subgroups are disadvantaged by the assessment.

Technical Advisory Committee Studies

In previous years, DRC and its partners have performed a number of validity and reliability studies. These have included studies of dimensionality, generalizability, and stability of the construct for different mixes of multiple choice and open-ended item types, dichotomous and partial credit scoring, for revisions of the content standards, and for on-line administration. DRC would be pleased to partner with PDE on any new studies that may arise in the future.

VII.J.8. Conducting Standard Setting Sessions

DRC will provide PDE with a detailed plan for establishing valid and legally defensible performance-level cut scores. The cut scores and performance levels will be based on the blueprints and content standards of the GCAs. The methodology used will be item mapping (a.k.a. *Bookmark*) for all assessments save for writing, which will use *Body of Work*

DRC has successfully conducted standard-setting meetings for the Commonwealth, as well as for many other clients in large-scale assessment (e.g., Alaska, Alabama, Idaho, Louisiana, and North Carolina). In each case, the standard-setting study was:

- Customized for the client.
- Presented to and accepted by the respective Technical Advisory Committees.
- Implemented.
- Documented.

- Presented to the school/state board for final approval.

DRC's experience in standard setting has shown that the following practices contribute to the validity of the process:

- Utilizing the performance descriptors to the greatest extent possible.
- Documenting the critical processes and elements of the study.
- Using the standard errors of the study to inform selection of the final cut points.
- Documenting the teaching experience and demographic characteristics of the participants.

Strict adherence to these practices and processes will ensure that the results from the meeting will yield legally defensible outcomes that meet or exceed industry standards.

DRC proposes that the standard setting for each GCA test occur after the first operational spring administration of that assessment because sample sizes are expected to be larger. For example, standard settings for the first three GCA tests (Algebra I, Algebra II) will take place after the Spring 2011 operational administration. The chart below illustrates estimated timelines for the GCA standard-setting meetings and subsequent State Board of Education approvals. Please refer to *Subheading VII.I.2* for more information on timelines.

GCA Course	Standard-Setting Meeting	State Board Approval (Estimated)
Algebra I, Algebra II, Geometry	June 7–10, 2011	June 16, 2011
Biology, Chemistry, English Composition, Literature, U.S. History, World History, Civics & Government	June 5–8, 2012	June 14, 2012

DRC's Proposed Standard Setting Team

DRC is pleased to propose **Dr. Melvin Webb** as the Standard-Setting Manager for all GCA standard-setting meetings. Mr. Webb received a BA in English from the University of North Carolina—Chapel Hill, an M.Ed. in Counseling/Student Personnel Work with minors in Educational Psychology and Public Administration from North Carolina State University, and an Ed.D. in Education Administration with minors in Research Design & Statistics and Public Policy Analysis from North Carolina State University.

Dr. Webb has worked in education since 1982 in a variety of roles and levels of responsibility, many of which apply directly to the role he will play in ensuring that all standards setting meetings for the GCAs produce valid results and legal

defensibility. For example, during his tenure at ACT Inc., he was Director of the NAEP Achievement Levels Setting Project, and subsequently led the effort to develop our nation's first performance level standards in reading, writing and mathematics. In addition, Dr. Webb was Senior Product Manager for Achievement & Aptitude Tests and Scoring & Reports for CTB/McGraw-Hill from 1996–2005.

Dr. Webb has also directly been involved with public schools in the Commonwealth; from 1993 to 1996, he served as the Director of the Office of Standards and Assessment for the School District of Philadelphia. He also served California public schools, working in the Sacramento City Unified School District from 2006–2008 as the Administrator for Assessment, Research & Evaluation.

Dr. Webb will be supported by **Mr. David Chayer, Vice President of Psychometric Services**. Mr. Chayer has extensive experience in Bookmark standard setting techniques as well as body of Work and has managed and provided training and facilitation for over a dozen large-scale meetings in both methods. These meetings have included projects for Pennsylvania, as well as Alaska, Idaho, Minnesota, and North Carolina. Mr. Chayer has performed and directed research, psychometric, and test development activities in norm-referenced, large-scale assessment and licensure/certification testing programs for both paper-and-pencil and computer-based testing.

Mr. Chayer holds a Master of Arts degree in Measurement, Educational Psychology, and a Bachelor of Arts degree in Statistics, with a minor in Philosophy, both from the University of Minnesota.

Project Description

For each GCA course, with the exception of Composition, DRC will use the item mapping (Bookmark) standard-setting method (Lewis, Mitzel, Green, & Patz, 1999) to set three cut points to get four performance levels (i.e., Below Basic, Basic, Proficient, Advanced). For Composition, DRC is proposing to use the Body of Work (BoW) method to set the cut points. The cut point that distinguishes the Basic from the Proficient performance level will be used to determine the Non-proficient/Proficient cut at each module. That is, a Proficient (passing) score on each module will represent the same level of performance as a Proficient score on the overall test.

This standard setting will include:

- Validation of the performance-level descriptors for each module for a GCA course.
- Appropriate training of standard-setting committee members in the Bookmark/BoW method for purposes of determining standards based on their knowledge, judgment, and use of consequential data.

- DRC staff who will lead and facilitate group discussions, including the processes for the standard setting. This will include the review and revision process as required for the performance-level descriptors for each course, module, and each performance level ensuring alignment with the tests and their current content standards.
- Tables needed to create the impact data used in the Bookmark/BoW standard-setting procedure.
- Recommended scaled score cut points for each exam within each content area.
- Technical Report of the process used to generate the recommended cut points.
- Technical documentation to PDE on the strategies and procedures used prior, during, and after the standard setting. Documentation of standard-setting data collected, results of analysis, achievement-level descriptors, and recommended standards based on committee judgment will be included.
- Executive summary containing the recommended cut scores from the panel group, along with the impact data provided to the group.

Project Tasks

Standard Setting Method

DRC recommends using the Bookmark procedure (Lewis, Mitzel, Green, & Patz 1999) for most GCA courses. The Bookmark methodology is consistent with previous standard settings and suitable for this project, as items can be reliably ordered by difficulty. In addition, the task required of the judges is considered less complex than the tasks required by other methods. Judges are asked to determine cut score(s) based on a difficulty scale and provide their judgments of items and the separation of one ability level from another.

The Bookmark standard-setting method has two components: the *ordered item booklet* (OIB), which presents test items in order of their scale (difficulty) locations as determined by item response theory (IRT) calibrations, and the *item map*, which contains both content and statistical information. The panelists record their individual judgments directly on the item map (in addition to placing a marker in the ordered item booklet).

In the OIB, the items are ordered by their scale (logit difficulty) locations, which also correspond to their ordering by classical item difficulties (p -values). The easiest item, based on scale score location, is placed in the front of the booklet, while the most difficult is placed at the back. This approach capitalizes on the desirable features of Rasch scaling techniques, which place both items and students on the same scale.

A primary feature of the Bookmark standard setting methodology is that panelists can make cut score judgments directly on the scale score metric, in the context of item content and grade-level expectations. The panelists place a bookmark in the

OIB at the point that divides the item content that a student at a given performance level should be able to answer from the item content that is too difficult. In this way, content difficulty is directly related to expectations for student performance.

Following several rounds of consideration, final cut scores are established by determining the median value of the cut scores (i.e., the median of the table cut scores, which are represented by the median of the respective panelists within the tables). Medians are generally preferred to means because they reduce the influence of extreme judgments, should any exist.

In addition to the OIB, participants will be provided with an item map and supplies, such as paper and adhesive notes. The item map is a table in which each row represents an item in the OIB, ordered in the same manner, with additional information as follows: (1) the scale location for the item, (2) the content categorization, (3) the source of the item (e.g., form and item number), and (4) space for panelists to record notes.

Standard Setting Panel

DRC will work with PDE in the recruitment of Pennsylvania educators for this process. DRC will contact (from the approved PDE list), assemble, and train the members for participation in this process.

The standard setting committee will be composed of a diverse group of teachers, including exceptional child (EC) specialists, English as second language (ESL) specialists, and curriculum specialists. They will be educators from Pennsylvania who have reviewed items in the past and have been recommended for, or have expressed interest in, the standard setting process. There will be a group of panelists, containing 18 members per course (3 tables, 6 panelists each). This group must be familiar with the subject matter (content), the population of students, the instructional environment, and other variables that might affect performance. DRC also recognizes the need to select panels that reflect the diversity in gender, ethnicity, and regional residence of Pennsylvania.

Materials

The materials that are central to the process include:

- The preliminary performance level descriptors, to define what students at each level should know and be able to do.
- An operational form of the test. While states vary in whether they provide participants with actual operational test booklets, DRC has found that it is useful for participants to see the items in exactly the same form as students saw them so that participants can experience the test in the same way that it is experienced by the students. DRC feels that the use of operational test booklets adds face validity to the standard setting process and allows the panelists to feel that their work is set within a real-world context.

- The OIB, to be used for placing the bookmarks.

The OIB will contain operational items from the assessment arranged in *difficulty* order. Each multiple-choice item will appear once in the booklet. For any item, all preceding items will be easier and all following items will be harder. Each open-ended item will appear as many times as there are score points. For each score point, there will be an example of actual student work that is typical of that score. They will be located at the scale score location that is implied by the score. This is the scale score for which the given score is the most probable outcome.

Bookmark Training

For success of the project, it is essential that the participants' understand the procedure. Each panelist will receive extensive training in a large-group setting prior to making any judgments. Panelists will receive an orientation to the Bookmark method and practice the mechanics of the process using a short "practice test" composed of non-secure training materials taken from a public source (e.g., released NAEP items). At the end of the training session, panelists will be asked to complete a short questionnaire regarding the adequacy of training activities. In addition, panelists could be asked to sign off on a "Readiness" form, indicating that they understand their task and are ready to begin with the bookmark placements. DRC will be happy to discuss PDE's preference upon contract award.

One important aspect of the training is the emphasis on the role of panelists to not make judgments about the wording or the difficulty of items. Rather, the role of the panelists is to carefully weigh the knowledge and skill levels necessary to have a 0.67 chance of correctly answering the questions.

The Bookmark Placement Task

Participants express their judgments of cutscores by placing a *bookmark* between the ordered items judged to represent the cut point. Three separate bookmarks are placed for each exam within a content area representing the cut points for the four performance levels. Training will emphasize the following points:

- The bookmark represents the panelist's judgment that all the subsequent items have less than a 67% likelihood of being answered by the student at the threshold of proficient.
- Bookmark placement should not be thought of as separating two items, but rather two groups of items. In other words, a placement should not hinge on distinctions drawn for adjacent items with similar locations. Rather, the collective locations of the group of items below the bookmark should be compared with the collective location of the group of items above the bookmark.
- Students with a scale score at a given location will have approximately a 0.67 probability of correctly responding to a multiple-choice item also at the location. These same students will have a higher probability of success on

easier items (before the bookmark placement) and a lower probability of success on harder items (after the bookmark placement).

Bookmark Process

The standard-setting process will involve three or more rounds of placing and reviewing the bookmark for each exam. There is no intent to reach a consensus; the panelists will be instructed to place their bookmark where they believe they should be, not where others in the group believe they should be. The first round will require each individual panelist to place the bookmark before any group discussion of the items.

Subsequent rounds will offer panelists the opportunity to revise their individual bookmark as additional information is provided to them. The additional feedback for Round 1 will include only the locations of the bookmarks for all panelists. This will give the panelists the opportunity to see how their decision compares to the other members of the group and to discuss the differences. Frequently, differences are traced to differing interpretations of the performance-level descriptors or different interpretations of the knowledge, skills, and competencies required by the items. During group discussions, the facilitators will encourage panelists to discuss and clarify the thinking that supports or refutes fellow panelists' interpretations of the performance-level descriptors and how they relate to their judgments.

DRC will work with PDE staff present at the standard setting to review the results of rounds prior to information being presented to the panelists.

Round 1

The first round of the bookmark process begins with a review of the ordered item booklets as part of a small group. Participants review each item, ordered in terms of difficulty, and are asked to determine and discuss what course knowledge, skills, and competencies are required to correctly respond to each item. In this way, items are directly compared, one to another, in terms of the content and skills that must be mastered for each successively more difficult item.

The Round 1 bookmark placements are made individually and discussion among group members is discouraged. This is intended to ensure that the Round 1 judgments are independent and to try to reduce the influence of other members' opinions or the opinion of a dominant group member. DRC believes that this round will provide the best estimate of the true inter-rater variability.

The panelists will be asked to place their *Proficient* bookmark first, at a place where they feel that a borderline *Proficient* student would have less than a 2/3 chance of answering correctly. Then, they will be asked to continue to go through the OIB and place their *Advanced* bookmark in the same manner. After that, they will be asked to go back to the start of the OIB to place their *Basic* bookmark.

At the completion of Round 1, the initial bookmarks defining the boundaries among performance levels from all panelists will be compiled by DRC staff and used to compute the group-level results.

Round 2

Panelists will begin Round 2 with an extensive discussion of their Round 1 ratings. This discussion typically begins at the small-group level, facilitated by the table leader. The discussion centers on what students should know at the two achievement levels. Results of the Round 1 judgments will be presented to the panelists at the beginning of Round 2, including a list of the Round 1 bookmark placements made by each panelist at each of the tables.

Following small-group discussion, a large-group discussion will be facilitated to incorporate more perspectives into Round 1 placements. Impact data, based on actual data from the GCA test administration, will be provided to help panelists frame the effects of their judgments.

After the large-group discussion, individual panelists will again review their original bookmark placements and make any adjustments in the bookmark placements they feel are appropriate. The judgments are entered into a spreadsheet program and the median cutscore is calculated for each small group and for the full panel. The latter is used to estimate the impact of the proposed standards.

All individual recommendations will then be collected, recorded, and analyzed. Feedback on the overall panel recommendation and the projected impact will be provided to the group as a whole.

Round 3

Panelists will begin Round 3 with extensive discussion of their Round 2 ratings. As in the previous rounds, the judgments from the prior round form the basis for the initial discussion. Each small group will discuss where they believe the cut should fall between the achievement levels and why.

Following small-group discussion, a large-group discussion (i.e., across tables) will be facilitated to incorporate additional perspectives into where the cut should be located. Impact data, from the live administration of the assessment, will be provided to help panelists frame the effects of their judgments.

Following the Round 3 large-group discussion, individual panelists will again review the placement of their bookmarks (in the OIB) and make the final bookmark placements. These judgments are once again entered into a spreadsheet program and the median cutscore is calculated for each small group, as well as for the full panel. The latter is used to estimate impact data.

All results for all rounds from the week of meetings will be summarized and recorded in a Technical Report for submission to PDE. Upon approval, DRC will generate the final scale score cut point for each module of each GCA course.

Use of Impact Data

The presentation of impact data is an important part of the process and adds an important aspect of consequential validity. Prior steps have relied on the panelists' understanding of the students, the curriculum, and the items to form judgments about appropriate performance standards. The impact data gives them information about how their proposed standards will affect students and schools. Specifically, the data presented will include the percentage of students who are placed in each of the two performance levels for a module. The question that they must now consider is if this is a fair and accurate description of the GCA students in the Commonwealth.

If the percentages in the levels are much higher or lower than the panelists expect and can defend, they may choose to alter placement of their bookmarks to better align with their expectation.

It is important to note that the reporting of results for the first two administrations of each GCA test (Fall 2010 and Spring 2011 for Wave 1 courses) will need to be delayed in order that impact data can be made available at the standard setting for panelists to consider. Similarly, results from the Fall 2011 and Spring 2012 administrations for the Wave 2 courses will be delayed due to standard-setting activities.

Computation of Standard Errors

Standard errors associated with the standard-setting process represent the likely range of recommendations that might result from an independent replication of the same process under the same conditions. There are two forms of error that are relevant to standard setting. One is the variability associated with the panelists' ratings and the second is the standard error of measurement of the test itself. The two types are described below, along with their respective roles in the process.

Standard Error of Panelists' Ratings

The standard error of the panelists' ratings applies to how well the recommended performance level is established. While standard deviations of the panelists' results will be computed for each round, the standard error of the group product will be based on the variability of the round one results. Round one is used because it represents the greatest degree of independence among the panelists. Later rounds tend to reflect more collaboration and discussion.

Standard of Error for Student Scale Scores

The other relevant standard error is commonly referred to as the *standard of error of measurement*. It differs from the standard error of the panelists' ratings in that it is a measure of the expected error of the person's scale score (ability estimate)

itself. The most informative use for the standard error of the scale score in standard setting processes is at the cut points, because they are the critical decision points for consequences and outcomes.

Evaluation of Standard Setting by Panelists

After the standard setting is complete, DRC will provide an opportunity for the panelists (judges) to evaluate the standard setting process in the form of an evaluation questionnaire. Results will be submitted in the technical documentation.

Body of Work

DRC will use the Body of Work method (Kingston, Kahl, Sweeny, & Bay, 2001) to set cut points for the GCA English Composition exam. DRC has successfully completed Body of Work standard settings in several states (including Pennsylvania for the PSSA) for Writing and various alternate assessments.

The Body of Work method is an examinee-centered approach in which panelists are asked to make a holistic evaluation of a student's collection of evidence, both multiple-choice and open-ended prompt responses, in order to place that examinee into the appropriate performance level. The evaluation materials used by the panelists will consist of a sample of examinees, with complete data sets from the Spring 2012 administration, who will be chosen randomly from each score point between just above chance level and the highest score. Panelists will not be informed of the examinee's original prompt scores, though they will be provided information about the examinee's multiple-choice responses including number correct.

The Body of Work method contains two phases: *rangefinding* and *pinpointing*. In the rangefinding phase, panelists will be presented with a sample of evaluation materials that cover the entire range of performance and asked to place each examinee into a performance level. DRC analysts will determine, for each panelist, the scale-score cut point resulting from their full-range classifications. In the pinpointing phase, panelists will then be presented with a collection of work samples consisting of the two examinees nearest each panelist's individual cut point, such that all panelists will review two examinee work samples for each panelist's individual cut point suggested in the rangefinding phase. All panelists review the same collection of response sets. This will allow the panelists to see how their classifications compare to others in the group. They will be given the opportunity to discuss the classifications and revise them individually if they desire. The intent is to ensure all panelists are comfortable with their own classifications and that all have the same understanding of the performance levels. Panelists will not be pressured to arrive at a consensus.

Training

Training will be conducted on the first morning of the meeting. Panelists will be told that they are to:

- Be responsible for all secure materials,
- Evaluate samples of student work and place in performance levels,
- Verify their individual placement for each round of judgments, and
- Participate in a discussion as a large group.

Training materials included:

- Performance Level Descriptors (PLD),
- Sample student work, and
- Rating form including the student's responses to multiple-choice questions.

Panelists will be told that the process will include iterations (rounds) of individual judgments, small group discussions and large group discussions, and opportunities to revise judgments. In addition, impacts will be presented impact data (i.e., percent of students in each performance level) based on the panelists' judgments; impacts would also, when appropriate, be presented for other courses.

Body of Work Process

Panelists will first engage in a rangefinding round, in which they are given one student response set (essay and multiple-choice responses) for each available score point after chance level. Panelists will be asked to classify each paper into the four performance levels, and also to mark whether they feel each paper is high, medium, or low in its assigned category.

In the pinpointing phase, DRC analysts will determine, for each panelist, the scale-score cut point resulting from their full-range classification. Panelists will then be presented with a collection of response sets consisting of the two examinees nearest each panelist's individual cut point, such that all panelists will review two examinee response sets for the cut point suggested in the rangefinding phase. All panelists will review the same collection of response sets; for example, if Panelist A proposes setting the cut point at point X and Panelist B proposes setting the cut point at point Y, all panelists would review the two examinee response sets closest to point X and the two examinee response sets closest to point Y. Panelists will be asked to make classification decisions by cut point; for the response sets surrounding the cut point, panelists will be asked to categorize each student's response into one of the four performance levels. Results will be shown to each group after both rangefinding and pinpointing.

Cut Points and Standard Errors

Cut points will be derived using logistic regression, which models the relationship between a continuous variable, such as a test score, and the probability of being in a binary category, such as being judged as Proficient. The form of the logistic regression equation is shown below:

$$\ln \frac{p}{1-p} = a + b \times x,$$

where a and b are the slope and intercept, respectively, of the logistic regression, and x is the score of interest. After each round, the judges' binary decisions will be used to estimate individual cut score estimates for each category in the logit metric. The medians of the individual estimates will then be used as overall estimates of the cut scores.

Computation of Standard Errors

Standard errors associated with this process represent the variability around the median of all (theoretical) judgments in the pool of Pennsylvania educators and stakeholders from which the panel will be chosen as a representative sample.

The calculations will be based on the standard error of the median. The standard error of the median, given a normal distribution, or a large sample, is approximately 25 percent larger than the standard error of the mean. Thus, the standard error of the mean is multiplied by a factor of 1.25 as a reasonable estimate of the standard error of the median.

To coincide with the goal of achieving the articulation of cutpoints across modules, the standard errors will be pooled across modules for English Composition.

Scaling and Transformations

Based on results of the Body of Work process, DRC analysts will derive linear equations, which will be used to convert student scores from the logit metric to the scale score metric for each course and module. The linear transformations will be chosen in consultation with the PDE to ensure the scale scores are meaningful and readily communicated.

Panelist Evaluation Survey

At the end of the standard setting process, panelists will be asked to complete an evaluation survey that reflects their level of satisfaction with the method, materials, training, process, individual and group judgments and recommendations, facilities, food, and use of time.

Technical Report

A draft of the technical documentation will be presented to PDE. At a minimum, this draft will include the following.

- History/purpose of the test
- Standard setting method
 - Name and description
 - Documentation from PDE on selection of judges

- Standard setting process
- Documentation on construction and implementation of materials used during the process
- Copies of non-secure materials used
- Training
- Panelist and group ratings for each round, including standard errors
- Final performance level descriptors
- Documentation of feedback received during the process
- Descriptive summary of the panelists' evaluation of the process and their confidence in their judgments
- Recommended cut score

Develop/Review and Validate Performance Level Descriptors for Each Course

DRC's Test Development staff will supervise the development, review, and validation of GCA performance level descriptors. The process to develop/review and validate Commonwealth of Pennsylvania draft general descriptions for performance levels for each test will include the following:

- What students in Pennsylvania should know and be able to do based on the content components of the GCA.
- How students would demonstrate this knowledge and skill based on the competency goals.
- The level of knowledge and skill necessary for each performance level.

In addition, the process will include making an initial determination as to:

- The necessary characteristics/concepts of achievement at each performance level.
- The categorization of the characteristics/concepts.
- The language is clear and easily understood.
- The description of the performance continuum.

DRC will submit descriptors to PDE for review. DRC will make PDE-requested edits and prepare the descriptors for review and use during the GCA standard setting sessions.

Performance Level Descriptor Training

DRC proposes that the committee members receive general training on how to review the performance level descriptors during the standard setting training. The training will be led by DRC content experts and will include a general overview of blueprints and content standards. Definition of key terms (e.g., course of study,

strands, competencies, objectives, performance level descriptors) will be provided.

Process

Following the general training session, Pennsylvania educators will be placed into committees by subject. Course-specific materials, including draft performance level descriptors, will be distributed and explained.

The purpose of each committee is three-fold:

- 1) First, participants will study the Pennsylvania standards, concepts, and competencies, as appropriate.
- 2) Next, DRC facilitators will present preliminary performance level descriptors to the participants. The committee members will be given time to completely familiarize themselves with each GCA assessment and the performance level descriptors for each assessment/module. They will review the descriptors, ensuring that each descriptor has an appropriate description of the performance level. They will also determine whether the descriptor is too demanding or too easy, if it is consistent with the expectation for the given GCA, or whether it should be modified.
- 3) Finally, participants review and revise the performance level descriptors to be used during the standard setting process.

DRC test development specialists will facilitate. The facilitators will ensure that a fair and orderly consensus process is followed, committee work products are adequately documented, and the process stays on schedule. The facilitators will also serve as resources, answering questions pertaining to the GCA assessment and the preliminary performance level descriptors that were prepared for review. DRC facilitators will summarize the results with suggested modifications listed, which will then be presented to the committee members. Open discussion will be encouraged until group consensus is met.

At the conclusion of the standard setting process, a sub-group of each course will review the performance level descriptors to determine if any final modifications are necessary. DRC will prepare and submit to PDE a final document with the performance level descriptors for each course. DRC will include all of the above in the Technical Report as well.

Security

All panel members will be required to sign an Original Work and Security Agreement form, and DRC will make sure that all members are aware that all work products will become the sole property of PDE and will be considered secure materials. DRC will provide panel members with sufficient copies of booklets with test items arranged in difficulty sequence.

First Administration Score Reporting

The reporting of results for the first administration (Dec. 2010 for Wave 1 courses and Dec. 2011 for Wave 2 courses) of each GCA test cannot occur until the performance levels have been established. The standard-setting meetings will be held following the first **spring** administration of the assessment. The reporting of results (with performance levels) for the first and second administration of each GCA test cannot occur until the performance levels have been established. Consequently, there will be a slight delay in reporting following these administrations to accommodate the standard-setting activity and the necessary approval process. This delay will not occur with subsequent assessments. For the first two administrations, since final results (including scaled scores and performance levels) will be delayed due to standard-setting activities, preliminary raw-score roster reports will be provided. For more information about the timelines, please refer to *Subheading VII.I.2*.

VII.J.9. Scoring Modules in each GCA

DRC understands that PDE wants to offer the GCAs in modules, so that students who do not pass the GCA may retake only the module(s) in which they were not successful. Our test design reflects this approach; please see *Subheading VII.B* for details and specifications regarding our test design for the GCAs. In addition, our reporting plan for the GCAs also supports this modular design; please see *Subheading VII.K* for full information regarding our proposed reporting system.

VII.J.10. Technical Report

DRC will produce an annual Technical Report following the spring assessments of the GCA. DRC believes its technical documents represent the best the industry has to offer. As stated in previous sections, the GCA Technical Report will serve as the primary vehicle for documenting reliability and validity evidence for the GCA. From the earliest stages of projects, DRC psychometricians are mindful of technical reporting and consider documentation needs continually. The same staff members who plan and conduct project analyses also prepare the associated technical documentation. As with project deliverables, the project's senior psychometrician oversees technical-report preparation. DRC's psychometric staff views the Technical Report as an essential component for enhancing the validity of the GCA assessments.

Contributions by other functional groups are managed with help of the DRC Project Management team. *Appendix 7* contains an example of one of DRC's Technical Reports for the Commonwealth.

DRC is continually seeking to improve processes. This includes preparation of technical documents. One example of this is the implementation of an internal review of technical documents by independent senior staff members. Cold reads by editors are used to eliminate errors associated with grammar and style.

DRC typically establishes a comprehensive core text for technical reports during the first year of a project. Thought is given to minimizing the amount of new text required yearly and to keeping text that requires modification to established locations. Whenever possible, program output is placed directly into technical documents to limit errors that might occur otherwise. Staff also use visual checks between statistics reported in technical documents and original program output.

In observance of the demand for quality assurance in the testing industry, DRC employs a **psychometric staff dedicated to Quality Assurance**. To ensure the accuracy and completeness of the GCA Technical Reports, Dr. Nhouyvanisvong and his team will work, alongside the Quality Assurance psychometricians and statistical analysts, checking for internal and external consistency and reasonableness. This, in conjunction with the tests and checks performed by the Software Quality Assurance Department, promises Technical Reports that will meet the highest standards.

Consistent with its current experience with PDE, DRC anticipates a solid working relationship with PDE and TAC. Both groups will be consulted regularly throughout the creation of the report. DRC foresees one edit and review cycle with PDE and the TAC resulting in revision and production of the final Technical Report.

The following pages include a sample table of contents for the GCA Technical Report. DRC will typically provide the assessment's purpose, test blueprint and test maps, test development procedures, reliability and validity results and graphics, scaling information, inter-rater agreement data, accommodations and testing of students with special needs, security information, administration details, scoring and equating procedures and results, standard setting results, reporting, and appropriate/inappropriate uses and interpretation of data. Appendices will include related materials, administrative regulations, state standards, sample items, committee rating forms, frequency/percentile distributions, state and system performance summaries by ethnic group, and other pertinent information in compliance with PDE.

The Graduation Competency Assessment (GCA)
Sample Table of Contents for Technical Report

Preface: Overview/Purpose

- Assessment Activities in the 2009 School Year

1.0 Background of Program

- Statewide Testing and Accountability
- Purpose of Program
- Organizations and groups involved

2.0 Test Development

- Overview of Assessment Test Specifications

- Test specifications for each subject

3.0 Item Development Process

- Analysis of Bank
 - Count of items
 - Count of retired items
 - Maintenance activities
- Test Blueprint
 - Test Maps
- Item Development
- Test Development Considerations
 - Universal Design
 - Depth of Knowledge
 - Test Item Writers
 - Readability of Test Items
 - Process of Item Construction
 - Item Content Review
 - Committee membership
 - Bias and Sensitivity Reviews
 - Committee membership
 - Item Data Review
 - Committee characteristics and selection process
 - Differential Item Functioning (DIF)
- Forms Construction

4.0 Test Administration

- Security
- Assessment Accommodations
 - Special Education
 - English Language Learners (ELL)
 - Braille
- Online Administration

5.0 Test Administration Procedures

- Test Sessions, Timing and Layout
- Shipping and Delivery Procedures
- Packaging and Delivery of Materials
- Materials Return
- Test Security Measures
- Assessment Accommodations

6.0 Processing and Scoring

- Receipt of Materials
- Scanning of Materials
- Scoring of Multiple-Choice Items
- Training
- Security

7.0 Scoring Open-Ended Items

- Test Scorers
- Writing prompt administration
- Open-ended Items
- Range-Finding and Anchor Papers
- Training Materials

- Inter-rater Agreement Data
- Additional Conditions for Scoring Writing
- Scorer Drift Study

8.0 Scaling, Calibration, and Item Analysis

- Rational
- The Rasch Measurement Model
- Scale Scores and Transformations
- Cut Points for Performance Levels
- Field test analysis

9.0 Equating

10.0 Reports

- Description of scores
- Reports
 - Scaling and Equating Procedures
 - Including estimates of error for equating
 - Differential Item Functioning (DIF)

11.0 Reliability

- Coefficient Alpha
- Internal Consistency
- Standard Errors of Measurement
- Subgroup Reliabilities
- Inter-judge agreement
- Decision Consistency at Performance Levels

12.0 Validity

- Content- and Curricular-Related Evidence
- Construct-Related Evidence
- Criterion-Related Evidence
- Correlations Among Strand Scores
- Factor Analysis of Strand Scores
- Validity Evidence for Different Student Populations
- Other validity studies as determined by PDE

13.0 Standard Setting/ Validation (as applicable)

- Performance Level Descriptors
- Bookmark Method
- Results

14.0 Other Studies

15.0 GCA Administrations

- Field Test
- Forms Construction

16.0 Spanish Language Assessments for Mathematics and Science

17.0 Quality-control Procedures

- Test Development
- Administration
- Scoring

- Psychometric Services
- Overall

18.0 Glossary of Terms

19.0 References

20.0 Appendices



VII.J.11. Technical Advisory Committee (TAC) Meetings

DRC understands that each statewide TAC meeting will be extended one day (to a period of three days) to allow for discussion of the GCA Program. As the current PSSA contractor, DRC is responsible for all administrative and logistical details for the TAC meetings so the addition of an extra meeting day can be handled in an efficient and streamlined manner. All appropriate GCA Program Team members will attend the TAC meetings. DRC will compensate TAC members for the additional TAC meeting day in the form of an honorarium at a rate of \$1,500 per member, per day. We acknowledge that three TAC meetings will be held each year.