TECHNICAL REPORT

for the
Pennsylvania
System of School Assessment

2006 Writing
Grades 5, 8, and 11

Provided by
Data Recognition Corporation

May 2007

Table of Contents

PREFACE: An Overview of Recent and Future Assessment

The period from 2003 through 2006 brought significant structural changes in the test blueprint for the Pennsylvania System of School Assessment (PSSA). These changes necessitated extensive test development and field-testing activity along with phased-in implementation in the operational assessment. Included in this process was the development and implementation of assessments in new or additional grade levels.

For reading and mathematics, content changes for grades 5, 8, and 11 were developed in 2003, field tested in spring 2004, and implemented in spring 2005. The *Technical Report for 2005 Reading and Mathematics* provides a description of test development activities, review of open-ended tasks and multiple-choice items, field testing, selection of items, statistical analysis of assessment data, reliability, validity, standard setting and other technical characteristics of the operational 2005 PSSA. Test development for the new grade levels of 4, 6, and 7 began in 2004, with field testing in 2005, and full implementation in 2006. Similarly, the *Technical Report for 2006 Reading and Mathematics: Grades 4, 6, and 7* provides a complete description of test development activities, item review, field testing, statistical analysis, item selection, and technical characteristics of the operational 2006 PSSA for these grade levels.

Changes in the writing assessment were designed to sharpen the focus on what is assessed with respect to Academic Standards 1.4 and 1.5. To support this effort, a shift in grade levels assessed was made, moving from grades 6 and 9 to grades 5 and 8, thereby aligning assessment to the end of elementary and middle school years as it is for reading and mathematics. The writing testing window was changed from Fall to February for grades 5 and 8, making it consistent with grade 11. Mode-specific scoring guidelines replaced domain scoring, and the introduction of stimulus-based passages and associated multiple-choice items measuring revising and editing contributed to a more valid conventions score. An account of the development of writing prompts and stimulus-based, multiple-choice items, review processes, field testing and item analysis, standard setting, and other technical characteristics of the operational 2006 PSSA may be found in the *Technical Report for 2006 Writing*.

To assist the reader in navigating through the year-to-year changes in all aspects of the PSSA, tables are presented along with explanatory text. Provided is an overview of the subject areas assessed, time of year the testing activity took place, and the type of testing that occurred (e.g., operational, field testing, grade 12 retest). [Please note that the grade 3 mathematics and reading assessment is not addressed in this technical report because CTB/McGraw-Hill, the scoring contractor for grade 3, is responsible for preparing that technical report.]

ASSESSMENT ACTIVITIES OCCURRING IN THE 2003–04 SCHOOL YEAR

Table P–1 outlines the operational assessment and field testing administered during the 2003–04 school year. (A spring operational assessment in mathematics and reading took place at grades 3, 5, 8, and 11).

As a result of new Assessment Anchor Content Standards (Assessment Anchors) developed by the Pennsylvania Department of Education (PDE) during 2003, new test items were developed (see Chapter Two of the *Technical Report for 2005 Reading and Mathematics*). Following the spring operational assessment, a separate, "standalone" field test of new items for grades 5, 8, and 11 was conducted. Note that grade 11 students also took an operational writing assessment in February, and grade 6 and grade 9 students participated in a fall writing assessment. Lastly, grade 12 students who as 11th graders in the preceding spring failed to attain at least the proficient level in any of the subject areas, were offered an opportunity to retest.

Table P–1. Operational Assessment and Field Testing During the 2003–04 School Year

Grade	Assessment Activity	Date
3	Operational Mathematics and Reading with embedded field test (conducted by CTB/McGraw-Hill)	April
4	No assessment	
5	Operational Mathematics and Reading	April
	Standalone field test in Mathematics and Reading	April/May
6	Operational Writing	October
7	No assessment	
8	Operational Mathematics and Reading	April
	Standalone field test in Mathematics and Reading	April/May
9	Operational Writing	October
11	Operational Mathematics and Reading	April
	Standalone field test in Mathematics and Reading	April/May
	Operational Writing	February
12	Retest opportunity for students who as grade 11 students in the spring of 2003 failed to reach at least the proficient level in mathematics, reading, or writing	October/ November

ASSESSMENT ACTIVITIES OCCURRING IN THE 2004–05 SCHOOL YEAR

Table P–2 displays the operational assessment and field testing that took place during the 2004–05 school year. The operational assessment at grades 5, 8, and 11 used items chosen from the Spring 2004 field test. This was the first operational assessment that reflected the Pennsylvania Assessment Anchors and Eligible Content. Fulfilling the No Child Left Behind Act of 2001 (NCLB) requirement that states must implement a test at grades 3 through 8, a major field test in mathematics and reading was administered at grades 4, 6, and 7. Item development for these new grade levels took place during 2004.

The grades 6 and 9 writing assessments were abandoned in favor of moving the writing assessment to grades 5 and 8. This accounts for the separate (standalone) field test at these grade levels. There was also a test administration change from October to February. The writing assessment also underwent changes to align the test to the Academic Standards for writing. New writing prompts and stimulus-based, multiple-choice items were also field tested at grade 11 as part of the operational assessment, hence the reference to an "embedded" field test. No assessment activity of any kind occurred at grade 9. As in fall 2003, the retest opportunity at grade 12 continued.

Table P–2. Operational Assessment and Field Testing During the 2004–05 School Year

Grade	Assessment Activity	Date
3	Operational Mathematics and Reading with embedded field test (conducted by CTB/McGraw-Hill)	April
4	Standalone field test for Mathematics and Reading	April
5	Operational Mathematics and Reading with embedded field test	April
	Standalone field test in Writing	February
6	Standalone field test for Mathematics and Reading	April
7	Standalone field test for Mathematics and Reading	April
8	Operational Mathematics and Reading with embedded field test	April
	Standalone field test in Writing	February
9	No assessment	
11	Operational Mathematics and Reading with embedded field test	April
	Operational Writing with embedded field test	February
12	Retest opportunity for students who as grade 11 students in the spring of 2004 failed to reach at least the proficient level in mathematics, reading, or writing	October/ November

ASSESSMENT ACTIVITIES OCCURRING IN THE 2005–06 SCHOOL YEAR

Table P–3 shows the assessment activities that occurred during the 2005–06 school year. Note that the reading and mathematics operational assessments ran consecutively from grades 3 through 8 and at grade 11. For grades 4, 6, and 7, it was the first year of an operational assessment. Field testing for mathematics and reading was embedded as part of the operational assessment at each grade level. At grade 3, the reference to field testing with items developed by DRC reflects the transition process of shifting the assessment from CTB/McGraw-Hill to DRC in 2007. As in previous years, the retest opportunity at grade 12 continued.

The first operational assessment for writing at grades 5 and 8 took place this year while the grade 11 writing assessment continued in the same February test window. New this year for all three grade levels, the operational writing assessment featured mode-specific scoring guidelines; stimulus-based, multiple-choice items; and a grade-specific emphasis shift in writing modes assessed. Since extensive field testing in February 2005 produced a pool of prompts for use over several years, no additional writing prompts were field tested in 2006. However, new multiple-choice items were field tested in the 2006 writing assessment.

Table P–3. Operational Assessment and Field Testing During the 2005–06 School Year

Grade	Assessment Activity	Date
3	Operational Mathematics and Reading with embedded field test of DRC-written items (conducted by CTB/McGraw-Hill)	April
4	Operational Mathematics and Reading with embedded field test	March
5	Operational Mathematics and Reading with embedded field test	March
	Operational Writing with embedded field test	February
6	Operational Mathematics and Reading with embedded field test	March
7	Operational Mathematics and Reading with embedded field test	March
8	Operational Mathematics and Reading with embedded field test	March
	Operational Writing with embedded field test	February
9	No assessment	
11	Operational Mathematics and Reading with embedded field test	March
	Operational Writing with embedded field test	February
12	Retest opportunity for students who as grade 11 students in the spring of 2005 failed to reach at least the proficient level in mathematics, reading, or writing	October/ November

ASSESSMENT ACTIVITIES PLANNED FOR THE 2006–07 SCHOOL YEAR

Table P–4 shows the assessment plan for the 2006–07 school year. Note that again the mathematics and reading assessments will be operational consecutively from grades 3 through 8 and at grade 11. For grades 4, 6, and 7, it will be the second year of an operational assessment and the first year in which these grade levels will be included in the AYP calculations. Field testing for mathematics and reading will continue to be embedded as part of the operational assessment at each grade level. This is the first year in which DRC will be responsible for the grade 3 assessment as the transition from CTB/McGraw-Hill is completed. As in the previous years, the retest opportunity at grade 12 will continue.

The operational assessment for writing at grades 5, 8, and 11 continues in the same February test window featuring the mode-specific scoring guidelines; stimulus-based, multiple-choice items; and a grade-specific emphasis in writing modes assessed, which were introduced in 2006. Since extensive field testing in February 2005 produced a pool of prompts for use over several years, no additional writing prompts will be field tested in 2007. However, new multiple-choice items will be field tested in the 2007 writing assessment.

Following the spring operational assessments, a separate, "standalone" field test in science is planned for grades 4, 8, and 11 with full implementation scheduled for 2008.

Table P–4. Operational Assessment and Field Testing During the 2006–07 School Year (Planned)

Grade	Assessment Activity	Date
3	Operational Mathematics and Reading with embedded field test	March
4	Operational Mathematics and Reading with embedded field test	March
	Standalone field test in Science	April/May
5	Operational Mathematics and Reading with embedded field test	March
	Operational Writing with embedded field test	February
6	Operational Mathematics and Reading with embedded field test	March
7	Operational Mathematics and Reading with embedded field test	March
8	Operational Mathematics and Reading with embedded field test	March
	Operational Writing with embedded field test	February
	Standalone field test in Science	April/May
9	No assessment	
11	Operational Mathematics and Reading with embedded field test	March
	Operational Writing with embedded field test	February
	Standalone field test in Science	April/May
12	Retest opportunity for students who as grade 11 students in the spring of 2006 failed to reach at least the proficient level in mathematics, reading, or writing	October/ November

Chapter One: Background of Pennsylvania System of School Assessment (PSSA)

This brief overview of assessment in Pennsylvania describes the original and subsequent legislative mandates, previous assessment programs, the history of the current program's development process, the program's intent and purpose, recent changes to the program, and the student population that participates in the assessments.

THE ORIGIN OF STATE ASSESSMENT IN PENNSYLVANIA

State assessment of student achievement came about as a result of legislation enacted in 1963. Generally known as the School District Reorganization Act (Act 299), the issue of whether large or small district size provided a better quality education led to the development of Section 299.1 of Act 299, which required the State Board of Education to

> . . . develop or cause to be developed an evaluation procedure designed to measure objectively the adequacy and efficiency of the educational program offered by the public schools of the Commonwealth . . . The evaluation procedure shall be so constructed and developed as to provide each school district with relevant comparative data to enable directors and administrators to more readily appraise the educational performance and to effectuate without delay the strengthening of the district's educational program. Tests developed . . . shall be used for the purpose of providing a uniform evaluation of each school district . . .

In response to the legislative mandate, the State Board of Education contracted with Educational Testing Service of Princeton, New Jersey, to engage in a two-year process of surveying and interviewing stakeholders in business, industry, education, and the general public as to what constituted a quality education. This led to the State Board adoption of *The Goals of Quality Education* in 1965. In 1967 the Department of Education formed an organizational unit along with staff to begin developing appropriate measures and engaging in extensive field testing during the 1967-68 and 1968-69 school years.

EDUCATIONAL QUALITY ASSESSMENT (EQA) PROGRAM

The first state assessment of students in Pennsylvania took place in the 1969-70 school year. Initially, state assessment was a purely school-based evaluation in the form of the *Educational Quality Assessment (EQA)* program, which reported grade 5 and 11 school-level results in ten goal areas. Grade 8 was added in 1974. Measuring both cognitive and non-cognitive areas, the program operated from 1970 through 1988. As the program evolved, a matrix sampling design was used in measuring and reporting school results in subject areas such as reading, language arts, mathematics, science, health, social studies, and analytical thinking. Initially, it operated as a voluntary program, but in 1974 it became mandatory on a cyclical basis.

TESTING FOR ESSENTIAL LEARNING AND LITERACY SKILLS (TELLS)

The next major revision in state assessment was the advent of the state's first mandated competency testing program, *Testing for Essential Learning and Literacy Skills (TELLS)* in the 1984–85 school year. The impetus for a statewide essential skills test evolved from an October 1983 document entitled *Turning the Tide: An Agenda for Excellence in Pennsylvania Public*

Schools. A two-pronged approach was advocated, calling for:

1. competency testing in grades 3, 5, and 8 as an "early warning system" to identify students with reading and mathematics difficulties, and
2. state-funded remedial instruction to provide needed additional help.

In response to this and other recommendations, the State Board of Education added *Chapter 3: Student Testing* to its regulations on June 14, 1984. It required all public school students in grades 3, 5, and 8 to be given criterion-referenced tests in reading and mathematics. The second part of the program, remedial instruction, was mandated by Act 93-1984, and required districts to provide remedial instruction programs to students identified by the tests given under the State Board regulation. Subsequently, funds were distributed to districts and intermediate units for this part of the program. The *TELLS* and *EQA* testing programs coexisted until the *EQA* was concluded in 1988. The *TELLS* program continued through the spring of 1991.

THE PENNSYLVANIA SYSTEM OF SCHOOL ASSESSMENT (PSSA)

The Pennsylvania System of School Assessment (PSSA) program was instituted in 1992. The PSSA returned to a school evaluation model with reporting at the school level only. Test administration took place in February/March, and school district participation was every third year based on the strategic planning cycle. Reading and mathematics were assessed at grades 5, 8, and 11; districts could choose to participate in the writing assessment at grades 6 and 9. State Board revisions to Chapter 5 in November 1994 brought major changes to the PSSA, beginning with the Spring 1995 assessment. These changes included:

1. all districts were required to participate in the reading and mathematics assessment each year,
2. student-level reports were generated in addition to school reports, and
3. the grades 6 and 9 writing assessment became mandatory on a three-year cycle corresponding to the district's strategic planning cycle.

PENNSYLVANIA ACADEMIC STANDARDS AND THE PSSA

A major structural change took place in test content with the State Board of Education's adoption of the Pennsylvania Academic Standards for Reading, Writing, Speaking and Listening, and Mathematics in January 1999 (Pennsylvania State Board of Education, 1999). The Academic Standards, which are part of *Chapter 4 Regulations on Academic Standards and Assessment*, detailed what students should know (knowledge) and be able to do (skills) at various grade levels. Subsequently, the State Board approved a set of criteria defining Advanced, Proficient, Basic, and Below Basic levels of performance. Reading and mathematics performance level results were reported at both the student and school levels for the 2000 PSSA. At that point, the PSSA became a standards-based, criterion-referenced assessment measuring student attainment of the academic standards while simultaneously determining the extent to which school programs enabled students to achieve proficiency of the standards.

ASSESSMENT ANCHOR CONTENT STANDARDS, CONTENT STRUCTURE, AND NEW GRADE LEVELS

Assessment in 2005 was marked by major structural changes in the PSSA. Assessment Anchor Content Standards (Assessment Anchors) developed during the previous school year to clarify content structure and improve articulation between assessment and instruction were implemented in terms of test design and reporting. At the same time, field-testing of mathematics and reading occurred at grades 4, 6, and 7. Year 3 calculations for AYP were conducted and reported.

The 2006 operational reading and mathematics assessment incorporated grades 4, 6, and 7 for the first time. The assessed grade levels for 2006 included grades 3 through 8 and 11. Year 4 calculations for AYP were conducted and reported for grades 5, 8, and 11. AYP calculations for grades 4, 6, and 7 will take place in 2007 when they are assessed for the second time.

PURPOSES OF THE PSSA

As outlined in Chapter 4 of the State Board Regulations, the purposes of the statewide assessment component of the PSSA are as follows:

1. Provide students, parents, educators, and citizens with an understanding of student and school performance.
2. Determine the degree to which school programs enable students to attain proficiency of academic standards.
3. Provide results to school districts (including charter schools) and Area Vocational Technical Schools (AVTSs) for consideration in the development of strategic plans.
4. Provide information to state policymakers, including the State Senate, the General Assembly, and the State Board, on how effective schools are in promoting and demonstrating student proficiency of academic standards.
5. Provide information to the general public on school performance.
6. Provide results to school districts (including charter schools and AVTSs) based upon the aggregate performance of all students, for students with an Individualized Education Program (IEP), and for those without an IEP.

The broad purpose of the state assessments is to provide information to teachers and schools to guide the improvement of curricula and instructional strategies to enable students to reach proficiency in the academic standards.

THE PENNSYLVANIA WRITING ASSESSMENT

In 1990 the state initiated an on-demand writing assessment in which students wrote an essay in response to a particular topic or prompt. Offered to school districts on a voluntary basis, the writing assessment consisted of three modes of writing: narrative, informational, and persuasive. The test administration for grades 6 and 9 used a matrix sampling design; nine prompts (three per mode) were administered to students within a school, although each student responded to just one randomly distributed prompt. Scoring was based on a six-point holistic scale. Student results were aggregated and reported at the school level only. In 1992 the writing assessment was incorporated as part of the PSSA. Beginning in 1995, districts were required to participate

in the writing assessment every third year in accordance with their strategic planning cycle. However, districts were also given the choice to participate more frequently. As a result, participation rose dramatically from the expected 167 districts (one-third) in any given year to 235 (47%) in 1995, 306 (61%) in 1996, 412 (82%) in 1997, 445 (89%) in 1998, and 449 (90%) in 1999.

With the advent of the Pennsylvania Academic Standards in 1999, major changes took place in the writing assessment, including alignment to the Academic Standards as well as changes in scoring method, prompts, testing date, and reporting. These changes, which are summarized below, were implemented in the 2000–01 school year and were followed by performance level reporting in the 2001–02 school year.

- The writing assessment became mandatory for all districts every year.
- Administration of the grades 6 and 9 writing assessment was changed from February to October.
- Scoring changed to a 4-point scale for each of five domains (focus, content, organization, style, and conventions).
- Prompts were different for grade 6 and grade 9 rather than being identical at the two grade levels.
- Within a grade level all students responded to two common prompts.
- The reporting model was greatly revised, and individual student reports were issued for the first time.
- A writing assessment for grade 11 was administered for the first time in February 2001.
- In 2002, performance levels were adopted for writing and implemented in the reporting of total writing results for the February grade 11 and Fall 2002 grades 6 and 9 writing assessment.

The 2006 PSSA operational writing assessment featured additional revisions in the writing assessment that included the following enhancements:

- A shift from grades 6 and 9 to grades 5 and 8, to provide better alignment to the end of elementary school and middle school.
- Grades 5 and 8 joined grade 11 in a February test window rather than the October window used previously for grades 6 and 9.
- Students responded to two writing prompts, which were evaluated in terms of (1) a mode-specific scoring guideline and (2) a conventions scoring guideline instead of the former domain scoring guideline.
- Stimulus-based revising/editing multiple-choice items were incorporated to provide a more reliable and valid measure of the conventions academic standard.

Chapter Two: New Test Development for Writing

The first PSSA operational writing test to include both multiple-choice items and writing prompts, aligned with the Academic Standards, was administered in the spring of 2006 to students in grades 5, 8, and 11. The items administered were field tested in the spring of 2005. The new writing assessment represents several fundamental changes over the previous operational assessment. The changes in the PSSA writing test include the development of writing tests at each grade level (5, 8, and 11) that are in alignment with the Academic Standard 1.4 (Types of Writing [Mode]) and Academic Standard 1.5 (Quality of Writing [Revising and Editing]). Below is a more detailed description of these changes and their rationale.

GRADE LEVEL

Starting with the 2006 operational test, students in grades 5, 8, and 11 are administered the writing test. The assessment of students in grades 5 and 8 (rather than grades 6 and 9) provides a clearer alignment with the end of elementary and middle school. Further, this allows schools to use information from the writing assessment to evaluate the effectiveness of their writing programs and to assess the needs of incoming students.

MULTIPLE-CHOICE ITEMS

Starting with the 2006 operational test, students at each grade level respond to twelve multiple-choice, stimulus-based revising/editing items. The use of multiple-choice items allows for a more reliable and valid measure of conventions (which include revising and editing) because it provides focused, predictable opportunities to assess students' skills in using conventions of language and writing.

WRITING PROMPT

Starting with the 2006 operational test, students at each grade level respond to two writing prompts. Students at grade 5 respond to prompts at two of the three modes (narrative, informational, persuasive). Each year, PDE selects two of the three modes for use in the test. Students at grades 8 and 11 respond to prompts at only the informational and persuasive modes. This change aligns with the expository forms of writing most often used in middle and high school curriculums, and it reflects the expectations for writing that occur in post-secondary classrooms and in the workplace.

WRITING ASSESSMENT MEASURES

In 1999, Pennsylvania adopted academic standards for writing *(Academic Standards for Reading, Writing, Speaking and Listening)* that describe what students should know and be able to do with the English language at a grade level. Within the framework of the new assessment, the writing prompts are measured under Academic Standards 1.4.A Narrative, 1.4.B Informational, and 1.4.C Persuasive, thus providing the responses to the eligible modes the prompts are designed to elicit. The writing prompts are also measured under Academic Standard 1.5.A–F Editing. The stimulus-based, multiple-choice items are measured under the Academic Standards 1.5.E Revising, and 1.5.F Editing.

OVERVIEW OF THE WRITING TEST

MULTIPLE-CHOICE ITEMS

Each multiple-choice item on the writing test is associated with an embedded passage containing errors. For operational testing in 2006, four multiple-choice items are associated with each passage. Multiple revising and editing instances are incorporated within each passage and require that a student demonstrate both passive and active revising and editing skills.

All multiple-choice items have four response options that include one single correct answer. The student is awarded one raw score point for choosing the correct response. Incorrect response choices, or distractors, typically represent some kind of misinterpretation or predisposition, unsound reasoning, or casual reading of the item and/or stimuli.

WRITING PROMPTS

At each assessed grade level, students respond to writing prompts developed to measure composition of writing as specified in the Academic Standards 1.4.A–C and further clarified in Academic Standards 1.5 A–G. The student response to a prompt requires approximately 60 minutes per prompt, though students are allowed more time to finish their responses if necessary. The writing prompts were field tested in 2005 with only one field test prompt being administered per student. Prompt modes were spiraled across the total number of available forms. Spiraling is accomplished by administering each student one of many available prompts in a sequential manner. For example, the first student receives Prompt 1, the second student Prompt 2, and so on until every prompt is administered. If there are more students than prompts, the sequence is repeated starting with the first prompt until every student is assigned a prompt. This process ensures that each prompt is administered to approximately equal and representative student populations in regard to demographics like gender, ethnicity, school size, and location in the state.

Beginning with the operational assessment in 2006, students in grade 5 respond to two prompts selected across three modes: narrative, informational, and persuasive. The narrative prompt can be story/fiction or personal narrative/recount, which aligns with Academic Standard 1.4.A. The informational prompt can be sequence (process analysis) or simple definition, which aligns with Academic Standard 1.4.B. The persuasive prompt can be problem/solution or evaluation, which aligns with Academic Standard 1.4.C. No writing prompts were field tested in 2006.

Beginning with the operational assessment in 2006, students in grade 8 respond to two prompts: informational and persuasive. The informational prompt can be sequence (process analysis), illustration, conceptual definition, cause/effect, classification, or compare/contrast, which aligns with Academic Standard 1.4.B. The persuasive prompt can be problem/solution or evaluation, which aligns with Academic Standard 1.4.C. No writing prompts were field tested in 2006.

Academic writing is the focus for the grade 11 PSSA writing assessment, including writing required for students who wish to pursue post-secondary educational and/or career opportunities. Beginning with the operational assessment in 2006, students in grade 11 respond to two prompts: informational and persuasive. The informational prompt can be advanced sequence (process analysis), illustration, definition, cause/effect, classification, or compare/contrast, which aligns with Academic Standard 1.4.B. The persuasive prompt can be problem/solution or evaluation, which aligns with Academic Standard 1.4.C. No writing prompts were field tested in 2006.

Beginning with the field test in 2005, the responses to writing prompts are scored twice using two different scoring guidelines developed especially for the PSSA. The first score is based on the application of a mode-specific scoring guideline, and the second score is based on the application of a conventions scoring guideline. The mode-specific scoring guideline is designed to evaluate first-draft, on-demand responses. It identifies the essential criteria for successfully responding to a particular mode of writing relating to the core areas of writing: focus, development of content, organization, and style. In contrast, the conventions scoring guideline measures the demonstrated level of control of sentence formation, grammar, usage, spelling, and punctuation. For more information on the application of the new scoring guidelines, see the current *Writing Item and Scoring Sampler*, available on the PDE Web site.

Copies of the scoring guidelines used to score the mode and the conventions are contained in Appendix A.

Chapter Three: Item/Test Development Process

Key activities in the development process of the 2006 operational writing test include initial item development, review of newly developed items, bias/fairness/sensitivity review, field testing of new items in 2005, field test item review with data, and final selection of items for the 2006 PSSA Writing Test. Table 3–1 provides a timeline of these major activities, which are described in this chapter and in *Chapter Five: Field Test Procedures*.

Table 3–1. General Timeline Associated with 2005 Field Test and 2006 Operational Assessment of Writing at Grades 5, 8, and 11

	Activities in the Item/Test Development Process	Timeframe
Field Test	Test Blueprint Developed / Finalized	February–May 2004
	Initial Item Development Conducted	May–September 2004
	PDE, with Pennsylvania Educators and Consultants, Reviewed Sample Items	June 2004
	Newly Developed Items Reviewed with PA Educators (New Item Review)	October 2004
	Items Reviewed for Bias, Fairness, and Sensitivity with PA Educators (Bias Review)	October 2004
	Field Test Forms Constructed	October–November 2004
	Test Materials Printed, Packaged, and Shipped	November 2004–January 2005
	Window for Test Administration	February 2005
	Rangefinding of Open-ended Field Test Items Conducted	March 2005
	Open-ended and Multiple-choice Field Test Items Scored	May–June 2005
	Field Test Results/Data Reviewed with PA Educators (Data Analysis)	July 2005
Core Operational	Operational Test Form Constructed	August–October 2005
	Test Materials Printed, Packaged, and Shipped	October 2005–January 2006
	Test Administration Window	February 2006
	Supplemental Rangefinding of Operational Open-ended Items Conducted	March 2006
	Operational Open-ended and Multiple-choice Items Scored	May–June 2006

TEST CONTENT BLUEPRINT FOR 2006

As indicated in Chapter One and Chapter Two, the PSSA is based on the Pennsylvania Academic Standards for Reading, Writing, Speaking, and Listening. The writing test specifically measures Academic Standards 1.4 (Types of Writing) and 1.5 (Quality of Writing). The Reading, Writing, Speaking and Listening Standards were designed to show what students should know and be able to do with the English language at a grade level. The Standards establish an outline for what can be assessed on the PSSA writing test and help to communicate the range of knowledge and skills from which the PSSA items would be designed.

The PSSA writing test for grades 5, 8, and 11 in 2006 followed this content blueprint and testing plan in order to reflect the Academic Standards.

2006 OPERATIONAL LAYOUT FOR WRITING

The PSSA operational layout was developed through the collaborative efforts of Data Recognition Corporation (DRC), the National Center for Improvement of Educational Assessment (NCIEA), and the Pennsylvania Department of Education. The layout was subsequently evaluated and approved by PDE. The writing test book is scannable and includes fields for student demographic data, stimuli (embedded error passages) linked to multiple-choice (MC) items, and writing prompts (WP). All MC items are worth 1 point. Responses to WP items receive a maximum of 4 points (on a scale of 1–4) for mode and also receive a maximum of 4 points (on a scale of 1–4) for conventions.

MULTIPLE-CHOICE ITEMS

Each test form contains a common set of operational items (i.e., each student takes an identical set of forms) along with matrix/embedded field test items. The matrix and embedded field test items are unique across form.

WRITING PROMPTS

Each test form contains two common operational writing prompts. These prompts are taken by all students at a grade level. The operational forms in 2006 do not contain matrix or embedded field test writing prompt.

FORMS

The 2006 PSSA Writing Test is comprised of ten forms. All of the forms contain the common items identical for all students and sets of unique ("matrix") items that fulfill several purposes. These purposes include

- expanding the total pool of items for school-level reporting.
- field testing new items.
- using items from the previous year's assessment for the purpose of equating/linking.

Tables 3–2 and 3–3 display the design for the writing test forms. The column entries for these tables denote:

A) No. of Core Revising and Editing (R&E) Stimulus-based MC Items per Form—Each multiple-choice item is associated with a stimulus-passage. This column provides the number of core (common) operational revising and editing multiple-choice items that appear per form. These items appear in every test form at a grade level.

B) No. of Matrix R&E Stimulus-based MC Items per Form—Each multiple-choice item is associated with a stimulus-passage. This column provides the number of matrix revising and editing multiple-choice items that appear per form. These items include linking MC items and field test MC items. Matrix items will be used for equating.

C) Total No. of R&E MC Items per Form—This column provides the total number of multiple-choice items that appear in one test form (Column A plus Column B).

D) No. of Pre-equated Core 4-pt. Writing Prompts (WPs) per Form—This column provides the number of operational core writing prompts. These prompts appear in every test form at a grade level. Pre-equating means that all of the prompts are on the same metric or scale before they are administered on an operational form.

E) Total No. of Forms—This column provides the total number of forms at a grade level. The values in this column are used as a multiplier to calculate figures in Column H and Column J.

F) Total No. of Core R&E Stimulus-based MC Items per 10 Forms—This column provides the total number of stimulus-based, multiple-choice items that appear in all core positions of all test forms. Since core items are identical across all forms, this number should equal the core figure provided in Column A.

G) Total No. of Matrix R&E Stimulus-based MC Items per 10 Forms (Linking & Embedded FT)—This column provides the total number of revising and editing multiple-choice items that appear in all matrix positions in all test forms. This figure is found by multiplying the number of matrix MC items shown in Column B by the total number of forms found in Column E. The matrix positions shown in Column H are further broken out into the number of Matrix Linking and the number of Embedded Field Test Matrix items. The total number of Matrix MC items is equal to the number of Matrix items and Matrix Linking items added to the number of Embedded Field Test Matrix items.

H) Total No. of R&E Stimulus-based MC Items (Core + FT) per 10 Forms—This column provides the total number of multiple-choice items that will appear in all MC positions in all forms regardless of role. This figure is found by adding the total in Column G to the total in Column H.

I) Total No. of Pre-equated Core 4-pt. Writing Prompts (WPs) per 10 Forms—This column provides the total number of writing prompts that appear in all forms. Since all writing prompts are core, the figure in Column K equals the figure in Column D.

Table 3–2. 2006 Writing Test Plan, per Operational Form per Grade

A	B	C	D	E
No. of Core R&E Stimulus-based MC Items per Form	No. of Matrix R&E Stimulus-based MC Items per Form	Total No. of R&E MC Items per Form	No. of Pre-equated Core 4-pt. Writing Prompts (WP) per Form	Total No. of Forms
12	8	20	2	10

Table 3–3. 2006 Writing Test Plan, per 10 Operational Forms per Grade

F	G	H	I
Total No. of Core R&E Stimulus-based MC Items per 10 Forms	Total No. of Matrix R&E Stimulus-based MC Items per 10 Forms (Linking & Embedded FT)	Total No. of R&E Stimulus-based MC Items (Core + FT) per 10 Forms	Total No. of Pre-equated Core 4-pt. Writing Prompts (WP) per 10 Forms
12	80 (16 & 64)	92	2

CORE POINTS

Since an individual student's score is based solely on the common, or core items, the total number of operational points is 100. The total score is obtained by combining the points from the core MC and WP portions of the test as displayed in Table 3–4.

Table 3–4. Maximum Eligible Core Points for Writing Prompts

Multiple-choice	**Writing Prompts**		**Totals**
	Conventions	**Mode**	
12	8	80	100
12 items × 1 point each (12×1)	2 items each worth a maximum of 4 points each (2×4)	2 items each worth a maximum of 4 points each. The raw score is then multiplied by 10. (2×4)×10	

TEST SESSIONS AND TIMING

The test window for the 2006 operational assessment was from February 13 through February 24, including make-ups. The writing assessment consists of three sections. Test administration requires each complete section to be scheduled as one assessment session, although schools are permitted to combine multiple sections as a single session. Administration guidelines stipulate that the sections be administered in the sequence in which they are printed in the test book. Table 3–5 outlines the assessment schedule and estimated times for each section.

Table 3–5. Writing—All Grades

Section	Contents	Suggested Time*
1	20 Multiple-choice	50 Minutes
2	1 Writing Prompt	60 Minutes
3	1 Writing Prompt	60 Minutes
	Total Time	150 Minutes

*These are approximate times. All students are entitled to extra time if needed. Students may request an extended assessment period if they indicate that they have not completed the task. Such requests are granted if the assessment administrator finds the request to be educationally valid.

REPORTING CATEGORIES AND POINT DISTRIBUTION

The writing assessment results will be reported in two categories:

- Composition
- Revising and Editing

Academic Standards A, B, and C are associated with Composition. Academic Standards E and F are associated with Revising and Editing. The distribution of core items into these two categories is shown in Table 3–6.

Table 3–6. Core Points Distribution

Reporting Category	*Composition*	*Revising and Editing*	
Academic Standards	*A–C*	*E and F*	
Multiple-choice Items	N/A	12	12
Writing Prompt 1	4 (Mode)	4 (Conventions)	8
Writing Prompt 2	4 (Mode)	4 (Conventions)	8
Raw Sub-total	8	20	28
Weighting Factor applied to Raw Score	*x10*	*x1*	
Total Possible Points	80	20	100

For more information concerning the process used to converting the operational layout into forms (form construction), see chapter 6. For more information about operational layout across forms and across years (form equivalency), see chapter 10.

TEST DEVELOPMENT CONSIDERATIONS

Achieving strong alignment of the items with the PSSA Academic Standards involves several components:

- grade-level appropriateness (reading/interest level, etc.)
- Webb's Depth of Knowledge (cognitive level, item/task level of complexity)
- estimated difficulty level
- relevancy of context
- rationale for distractors
- style
- accuracy
- correct terminology

The inclusion of multiple components such as these greatly enhances the comprehensiveness and utility of alignment (Bhola, Impara & Buckendahl, 2003). *The Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) and the *Principles of Universal Design* (Thompson, Johnstone & Thurlow, 2002) guided the development process. In addition, DRC's *Bias, Fairness, and Sensitivity Guidelines*[*] were used for developing items free of issues of bias, fairness, and sensitivity. All items were reviewed for fairness by bias and sensitivity committees and for content by Pennsylvania educators and field specialists.

BIAS, FAIRNESS, AND SENSITIVITY

At every stage of the item and test development process, DRC employs procedures that are designed to ensure that items and tests meet Standard 7.4 of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999).

> *Standard 7.4: Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain.*

[*] A printed copy of DRC's current edition of the *Bias, Fairness, and Sensitivity Guidelines* may be obtained by writing to:

ATTN: *Bias, Fairness, and Sensitivity Guidelines* Document Request
Test Development
Data Recognition Corporation
13490 Bass Lake Road
Maple Grove, MN 55311

In meeting Standard 7.4, DRC employs a series of internal quality steps. DRC provides specific training for our test developers, item writers, and reviewers on how to write, review, revise, and edit items for issues of bias, fairness, and sensitivity (as well as for technical quality). Our training also includes an awareness of and sensitivity to issues of cultural diversity. In addition to providing *internal* training in reviewing items in order to eliminate potential bias, DRC also provides *external* training to the review panels of minority experts, teachers, and other stakeholders.

DRC's guidelines for bias, fairness, and sensitivity includes instruction concerning how to eliminate language, symbols, words, phrases, and content that might be considered offensive by members of racial, ethnic, gender, or other groups. Areas of bias that are specifically targeted include, but are not limited to: stereotyping, gender, regional/geographic, ethnic/cultural, socioeconomic/class, religious, experiential, and biases against a particular age group (ageism) and against persons with disabilities. DRC catalogues topics that should be avoided, and maintains balance in gender and ethnic emphasis within the pool of available items and passages.

UNIVERSAL DESIGN

The principles of Universal Design were incorporated throughout the item development process to allow participation of the widest possible range of students in the PSSA. The following checklist was used as a guideline:

1. Items measure what they are intended to measure.
2. Items respect the diversity of the assessment population.
3. Items have a clear format for text.
4. Items have concise and readable text.
5. Items allow changes to format, such as Braille, without changing meaning or difficulty.
6. The arrangement of the items on the test has an overall appearance that is clean and well organized.

A more extensive description of the application of Universal Design principles is described in Chapter Four.

DEPTH OF KNOWLEDGE (DOK)

Important in statewide assessment is the alignment between the overall assessment system and the state's standards. A Depth of Knowledge (cognitive complexity) methodology developed by Webb (1999) offers a comprehensive model that can be applied to a wide variety of contexts. With regard to the alignment between standards statements and the assessment instruments, Webb's criteria include five categories, one dealing with content. Within the content category is a useful set of levels for evaluating DOK. According to Webb (1999, p.7–8) "Depth of Knowledge consistency between standards and assessments indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards." The four levels of cognitive complexity (Depth of Knowledge) are:

- **Level 1: Recall**—the student can recall information and facts.

- **Level 2: Skill/Concept**—the student can use information and facts in new situations.
- **Level 3: Strategic Thinking**—the student can use reason and strategic thinking to develop a plan.
- **Level 4: Extended Thinking**—the student can use extended thinking and investigation to solve a problem.

DOK levels were incorporated in the item writing and review process, and items were coded with respect to the level they represented. For the writing assessment, multiple-choice items are written to DOK levels 1 and 2. Students will either recall information they have learned (level 1) or apply learned information to a new context (level 2). For example, an item that asks students to identify a capitalization error (What is the capitalization rule for this word?) would be considered DOK level 1 because it requires learned facts. An item that asks students to insert a new sentence into an existing passage would be considered DOK level 2 because it requires more than mere recollection to arrive at the correct answer. The writing prompts are considered DOK level 3 because the student must create a unique piece of writing.

PROCESS OF ITEM CONSTRUCTION

As part of the item construction process, each item was reviewed by content specialists and editors at DRC, at WestEd, or at both companies (depending on the grade). Content specialists and editors evaluated each item to make sure that it measured the intended eligible content and/or assessment anchor (Assessment Anchor Content Standards). They also assessed each item to make certain that it was appropriate to the intended grade and that it provided and cued only one correct answer. In addition, the difficulty level, Depth of Knowledge, graphics, language demand, and distractors were also evaluated. Other elements considered in this process include, but is not limited to: Universal Design, bias, source of challenge, grammar/punctuation, and PSSA style.

A flow chart summarizing the item and test development processes used appears in Appendix B. Additional details about the process is discussed below.

SAMPLE ITEM REVIEW: JUNE 2004

Before training item writers to construct items, passages, and prompts for the new Pennsylvania writing test, DRC assessment and content experts developed a draft item construction orientation manual specifically for the PSSA writing assessment. This manual provided guidelines for the types of items and the character of the items to be developed for the assessment. In conjunction with this manual, DRC prepared a series of sample passages, prompts, and items that illustrated the initial view of what the future test items might look like.

A group of Pennsylvania educators was convened in Harrisburg on June 29 and 30, 2004, to review the proposed training materials and the samples of proposed item types. The Sample Item Review committee consisted of Pennsylvania teachers and subject-area supervisors from school districts throughout the Commonwealth of Pennsylvania, including some with post-secondary university affiliations. During this review, DRC received valuable feedback on the range and character of the items that Pennsylvania educators expected on the upcoming assessment. Committee members suggested revisions and made recommendations for reclassification of items. The committee also reviewed the items for adherence to the principles

of Universal Design, including language demand and issues of bias, fairness, and sensitivity.

Following this review, DRC consulted with PDE regarding the suggestions made by the committee members and made agreed-upon revisions to the training materials. This manual, along with DRC's standard item writing manual, were then used to train item writers to construct items for the Pennsylvania assessment.

TEST ITEM WRITERS AND TRAINING IN ITEM WRITING: CONSTRUCTING PROMPTS, PASSAGES, AND MULTIPLE-CHOICE ITEMS

The prompts, embedded-error passages, and multiple-choice items were developed by DRC ELA/writing test development specialists, scoring directors, and writers who have experience writing prompts and items for English language arts and writing assessments. Qualified writers were professionals with language arts classroom experience or writers who demonstrated appropriate grade-level content knowledge. Writers attended a one-day training workshop and were provided with a detailed instruction manual. As they wrote and revised their passages and items, writers also received personalized feedback from DRC test development content specialists.

Before developing items for the PSSA, the item writers were also trained in the following:

- Pennsylvania Academic Standards
- Webb's Four Levels of Cognitive Complexity: Recall, Basic Application of Skill/Concept, Strategic Thinking, and Extended Thinking
- General scoring guidelines
- Specific and general guidelines for item writing
- Bias, fairness, and sensitivity
- Principles of Universal Design
- Item quality technical style guidelines
- Reference information
- Sample items

In addition to the above, the training for passage, prompt, and item writing included guidelines on appropriate length, grade-level interest, and grade-level vocabulary. The training for multiple-choice items also included guidelines on proportionate distribution of items addressing each standard at each grade level, and general item construction guidelines to meet PDE's stated preferences. (For example, writers were told to use the phrase "incomplete sentence" rather than "sentence fragment.") The training for prompts also included special emphasis on Universal Design, clarity, validity, reliability, structure, format, interest, content, and vocabulary.

DRC sought and obtained from its item writers about twice as many passages and items as were needed to actually appear on the field test. The extra items allow future review committees to reject items and also allow DRC to select only the best items to move forward at each stage of development.

To ensure that the items were sufficient in number and adequately distributed across

subcategories and levels of difficulty, writers were assigned a specific number of items to create and attach to each passage.

Since all passages were written on commission, the passages were purchased outright, eliminating the need to seek costly permissions later when the passages reached publication.

Accepted passages and items then underwent an internal review by test development content specialists, content editors, and testing experts to judge their merit with regard to the following criteria:

- Passages and prompts have interest value for students.
- Passages and prompts as a whole demonstrate topical variety.
- Passages, prompts, and items are grade appropriate in terms of vocabulary, length, and language characteristics.
- Passages, prompts, and items are free of bias, fairness, and sensitivity issues.
- Passages contain common, grade-appropriate errors.
- Prompts and items measure only one standard.
- Prompts and items are clear, concise, and parallel in structure.
- Items are, as much as possible, passage dependent.
- Items provide for a range of difficulty.
- Prompts are rich enough to elicit measurable responses.

Once through the internal DRC review process, those passages, prompts, and items deemed potentially acceptable were then reviewed and approved by two PDE-sponsored committees: the New Item Review Writing Content Committee and the Bias, Fairness, and Sensitivity Committee.

PASSAGE, PROMPT, AND ITEM CONTENT REVIEW: OCTOBER 2004

Before field testing, all newly developed test items were submitted to content committees for review. The content committees consisted of Pennsylvania teachers and subject-area supervisors from school districts throughout the Commonwealth of Pennsylvania, some with post-secondary university affiliations. The primary responsibility of the content committee was to evaluate items with regard to quality and content classification, including grade-level appropriateness, estimated difficulty, DOK, and source of challenge. The committee members suggested revisions and made recommendations for reclassification of items. In some cases an item was deleted, and the committee suggested a replacement item and/or reviewed a suggested replacement item provided by the facilitators. The committee also reviewed the items for adherence to the principles of Universal Design, including language demand and issues of bias, fairness, and sensitivity.

The content review was held October 4–7, 2004. Committee members were PDE-approved. PDE internal staff members and DRC testing experts were also in attendance. The meeting commenced with an overview of the test development process by Patricia McDivitt, Vice President of Test Development (DRC). Ms. McDivitt also provided training on the procedures

and forms to be used for item content review.

DRC assessment specialists in writing facilitated the reviews. Committee members, grouped by grade level, worked through and reviewed the passages, prompts, and items for quality and content, as well as for the following categories designated on a generic Item Review Form, which may be found in Appendix C:

- Content Alignment
- Rigor Level Alignment
- Technical Design

Within these three areas, reviewers checked the standard being assessed, the grade level appropriateness, DOK, source of challenge issues, and the validity of the answer options. DRC assessment specialists recorded focused information on this form and on the item cards themselves.

The committee members then assigned each item a status: Accept or Edit. If a passage or item was revised, committee members agreed on the revision. All comments were recorded, collected, and filed.

Security during the meetings was addressed by adhering to a strict set of procedures. Passages and items in numbered binders were distributed for committee review and signed in and out by each member on a daily basis. All attendees, with the exception of PDE staff, were required to sign a Confidentiality Agreement. Secure materials that did not need to be retained after the meetings were deposited in secure barrels, and their contents were shredded.

BIAS, FAIRNESS, AND SENSITIVITY REVIEWS

Before field testing, all newly developed writing test items for grades 5, 8, and 11 were submitted to a Bias, Fairness, and Sensitivity Committee for review. This took place on October 5–8, 2004. The committee's primary responsibility was to evaluate items as to acceptability with regard to bias, fairness, and sensitivity issues. They made recommendations for changes or deletion of items to address bias, fairness, and/or sensitivity. Included in the review were proposed writing passages used as stimuli for the multiple-choice items. An expert multi-ethnic committee composed of men and women was trained by a DRC test development director to review items for bias, fairness, and sensitivity issues. Training materials included a manual developed by DRC (DRC, 2003). One committee member had expertise with special needs students. Another member worked for the Pennsylvania Department of Education in the curriculum department. Most of the writing items were read by all members, and some were read by a cross-section of committee members. Each member noted bias, fairness, and/or sensitivity comments on tracking sheets and on the item, if needed, for clarification. All comments were then compiled and decisions on the actions to be taken were made by the DRC writing test development specialists in consultation with PDE. This review followed the same security procedures as outlined above, except that the materials were locked and stored at the DRC offices in Harrisburg while in use and then shredded at the meeting's adjournment.

ITEM AUTHORING AND TRACKING

Initially, items are prepared on PSSA Item Cards and used for preliminary sorting and review. Although very similar, the PSSA Item Card for multiple-choice items differs from the PSSA Item Card for passages in that the former has a location at the bottom of the card for comments regarding the distractors. Blank examples of these two cards are shown in Appendix D. In both instances, a column against the right margin provides for codes to identify the subject area, grade, content categories, passage information, item type, DOK, estimated difficulty, and answer key (MC items).

All items undergoing field testing were entered into the DRC Item Viewer and Authoring Network ™ (IVAN), which is a comprehensive, secure, online item banking system. It accommodates item writing, item viewing and reviewing, and item tracking and versioning. IVAN manages the transition of an item from its developmental stage to its approval for use within a test form. The system supports an extensive item history that includes item usage within a form, item-level notes, content categories, and subcategories, item statistics from both classical and Rasch item analyses, and classifications derived from analyses of differential item functioning (DIF). A sample IVAN Item Card is presented in Appendix D.

Chapter Four: Universal Design Procedures Applied in the PSSA Test Development Process

Universally designed assessments allow participation of the widest possible range of students and contribute to valid inferences about participating students. Principles of Universal Design are based on the premise that each child in school is a part of the population to be tested, and that testing results should not be affected by disability, gender, or race (Thompson, Johnstone & Thurlow, 2002). At every stage of the item and test development process, including the 2005 writing field test and 2006 operational test, procedures were employed to ensure that items and subsequent tests were designed and developed using the elements of universally designed assessments that were developed by the National Center on Educational Outcomes (NCEO).

Federal legislation addresses the need for universally designed assessments. The *No Child Left Behind Act* (Elementary and Secondary Education Act) requires that each state must "provide for the participation in [statewide] assessments of all students" [Section 1111(b)(3)(C)(ix)(l)]. Both Title 1 and IDEA regulations call for universally designed assessments that are accessible and valid for all students, including students with disabilities and students with limited English proficiency. The benefits of universally designed assessments not only apply to these groups of students, but to all individuals with wide-ranging characteristics.

Committees involved in content and bias reviews included members familiar with the unique needs of students with disabilities and students with limited English proficiency.

What follows are the Universal Design guidelines followed during all stages of the item development process for the PSSA writing field test and operational test.

ELEMENTS OF UNIVERSALLY DESIGNED ASSESSMENTS

After a review of research relevant to the assessment development process and the principles of Universal Design (Center for Universal Design, 1997), NCEO has produced seven elements of Universal Design as they apply to assessments (Thompson, Johnstone & Thurlow, 2002). These elements guided PSSA item development.

- **Inclusive Assessment Population**
  The PSSA target population includes all students at the assessed grades attending Commonwealth schools. For state, district, and school accountability purposes, the target population includes every student except those who will participate in accountability through an alternate assessment.

- **Precisely Defined Constructs**
  An important function of well-designed assessments is that they actually measure what they are intended to measure. The Pennsylvania Academic Standards provide clear descriptions of the constructs to be measured by the PSSA at an assessed grade level. Universally designed assessments must remove all non-construct-oriented cognitive, sensory, emotional, and physical barriers.

- **Accessible, Non-biased Items**
  DRC conducted both internal and external reviews of items and test specifications to ensure that they did not create barriers because of lack of sensitivity to disability, culture, or other subgroups. Item and test specifications were developed by a team of individuals who understand the varied characteristics of items that might create difficulties for any group of students. Accessibility is incorporated as a primary dimension of test specifications, so that accessibility was woven into the fabric of the test rather than being added after the fact.

- **Amenable to Accommodations**
  Even though items on universally designed assessments are accessible for most students, there are some students who continue to need accommodations. This essential element of universally designed assessment requires that the test is compatible with accommodations and a variety of widely used adaptive equipment and assistive technology. (See Assessment Accommodations on page 25.)

- **Simple, Clear, and Intuitive Instructions and Procedures**
  Assessment instructions should be easy to understand, regardless of a student's experience, knowledge, language skills, or current concentration level. Knowledge questions that are posed within complex language can invalidate the test if students cannot understand how they are expected to respond to a question. To meet this guideline, directions and questions were prepared in simple, clear, and understandable language that underwent multiple reviews.

- **Maximum Readability and Comprehensibility**
  A variety of guidelines exist to ensure that text is maximally readable and comprehensible. Readability and comprehensibility are affected by many characteristics, including student background, sentence difficulty, text organization, and others. All of these features were considered as item text was developed.

  Plain language is a concept now being highlighted in research on assessments. Plain language has been defined as language that is straightforward and concise. The following strategies for editing text to produce plain language were used during the editing process of the new PSSA items.

  - Reduce excessive length
  - Use common words (unless necessary in context of the measurement of the item)
  - Avoid ambiguous words
  - Avoid inconsistent naming and graphic conventions
  - Avoid unclear signals about how to direct attention

- **Maximum Legibility**
  Legibility is the physical appearance of text, the way that the shapes of letters and numbers enable people to read text easily. Bias results when tests contain physical features that interfere with a student's focus on or understanding of the constructs that test items are intended to assess. A style guide (DRC, 2004-2006) was used, with PDE approval, that included dimensions of style consistent with Universal Design.

GUIDELINES FOR UNIVERSALLY DESIGNED ITEMS

All test items written and reviewed adhered closely to the following guidelines for Universal Design. Item writers and reviewers used a checklist during the item development process to ensure that each aspect was attended to.

1. **Items measure what they are intended to measure.** Item writing training included ensuring that writers and reviewers had a clear understanding of Pennsylvania's Academic Standards. During all phases of test development, items were presented with content standard information to ensure that each item reflected the intended standard. Careful consideration of the content standards was important in determining which skills involved in responding to an item were extraneous and which were relevant to what was being tested.
2. **Items respect the diversity of the assessment population.** To develop items that avoid content that might unfairly advantage or disadvantage any student subgroup, item writers, test developers, and reviewers were trained to write and review items for issues of bias, fairness, and sensitivity. Training also included an awareness of and sensitivity to issues of cultural and regional diversity.
3. **Items have a clear format for text.** Decisions about how items are presented to students must allow for maximum readability for all students. Appropriate typefaces and sizes were used with minimal use of italics, which is far less legible and is read considerably more slowly than standard typefaces. Captions, footnotes, keys, and legends were at least a 12-point size. Legibility was enhanced by sufficient spacing between letters, words, and lines. Blank space around paragraphs and ragged right margins were used.
4. **Items have concise and readable text.** Linguistic demands of stimuli and items can interfere with a student's ability to demonstrate knowledge of the construct being assessed. During item writing and review, the following guidelines were used.
    - Simple, clear, commonly used words were used whenever possible.
    - Extraneous text was omitted.
    - Vocabulary and sentence complexity were appropriate for the grade level assessed.
    - Technical terms and abbreviations were used only if related to the standard being measured.
    - Definitions and examples were clear and understandable.
    - Idioms were avoided unless idiomatic speech was being assessed.
    - The questions to be answered were clearly identifiable.
5. **Items allow changes to format without changing meaning or difficulty.** A Large Print and a Braille version of the PSSA were available at each assessed grade. Specific accommodations were permitted such as signing to a student, the use of oral presentation under specified conditions, and the use of various assistive technologies.
6. **The test has an overall appearance that is clean and organized.** Text that may not be

necessary and may be potentially distracting to students was avoided. Also avoided were purely decorative features that did not serve a purpose. Information was organized in a manner consistent with an academic English framework with a left-right, top-bottom flow.

ITEM DEVELOPMENT

DRC works closely with the Pennsylvania Department of Education to help ensure that PSSA tests comply with nationally recognized Principles of Universal Design. We support the implementation of accommodations on large-scale statewide assessments for students with disabilities. In addition to the Principles of Universal Design as described in the Pennsylvania Technical Report, DRC applys the standards for test accessibility as described in *Tests Access: Making Tests Accessible for Students with Visual Impairments— A Guide for Test Publishers, and State Assessment Personnel* (Allman, 2004). To this end, we embrace the following precepts:

- Test directions are carefully worded to allow for alternate responses to writing prompts.
- During item and bias reviews, test committee members are made aware of the Principles of Universal Design and of issues that may adversely affect students with disabilities with the goal of ensuring that PSSA tests are bias free for all students.
- DRC special education content specialists review items with the goal of ensuring that they are universally designed and accessible.
- With the goal of ensuring that the PSSA tests are accessible to the widest range of diverse student populations, PDE instructs DRC to limit item types that are difficult to format in Braille, and that may become distorted when published in large print. DRC is instructed to limit the following on the PSSA.
    - Unnecessary boxes and framing of text, unless enclosing the text provides necessary context for the student; use of italics (limited to only when it is absolutely necessary).

ITEM FORMATTING

DRC formats PSSA tests to maximize accessibility for all students by using text that is in a point size and font style that is easily readable. We limit spacing and number of items per page so that there is sufficient white space on each page. Whenever possible, we ensure that tables are positioned on the page with the associated test items. We use high contrast for text and background where possible to convey pertinent information. Tests are published on dull-finish paper to avoid the glare encountered on glossy paper. DRC pays close attention to the binding of the PSSA test booklets to ensure that they lie flat for two-page viewing and ease of reading and handling.

DRC ensures consistency across PSSA assessments by following these Principles of Universal Design:

- High contrast and clarity is used to convey detailed information.
- Typically, shading is avoided; when necessary for content purposes, 10% screens are used as the standard.
- Overlaid print is avoided.

- Tables are clearly labeled with titles and with short descriptions where applicable.
- Only relevant information is included in tables.
- Symbols used are meaningful and provide reasonable representations of the topic they depict.

ASSESSMENT ACCOMMODATIONS

While universally designed assessments provide for participation of the widest range of students, many students require accommodations to participate in the regular assessment. The intent of providing accommodations for students is to ensure that they are not unfairly disadvantaged during testing and that the accommodations used during instruction, if appropriate, are made available as students take the test. The literature related to assessment accommodations is still evolving and often focuses on state policies regulating accommodations rather than on providing empirical data that supports the reliability and validity of the use of accommodations. On a yearly basis, the Pennsylvania Department of Education examines accommodations policies and current research to ensure that valid, acceptable accommodations are available for students. An accommodations manual entitled *2006 Accommodations Guidelines* (PDE, December 2005) was developed for use with the 2006 PSSA.

Chapter Five: Field-Test Procedures

FIELD-TEST FORMS CONSTRUCTION

Field-test forms construction took place from October through December of 2004, after the Item Content Review. All items without an "Accepted" status were revised according to committee recommendation and then approved by PDE. DRC designed the field test format and received PDE approval. Thirty forms, labeled 01–30, were then constructed for each grade, with items distributed across the assessable Academic Standards. Each form contained ten multiple-choice field test items and one field test prompt. Grade 11 forms also contained two operational prompts.

GRADES 5 AND 8 TEST PLAN: STANDALONE FIELD TEST

Each student taking the standalone field test was administered one writing prompt and two short stimulus passages, with five stimulus-based revising and editing multiple-choice items per passage. The stimulus passage, with multiple-choice items, was administered before the writing prompt. The estimated administration time for the grades 5 and 8 standalone field test was approximately 60–75 minutes. Table 5–1 provides a summary of the spring 2005 grades 5 and 8 standalone writing field test.

Table 5–1. Grades 5 and 8 Writing Field Test Plan, Including Writing Prompts and Stimulus-Based Multiple-Choice Items

Grade	No. of WP Needed for Field Test in Spring 2005	No. of WP Needed for Op. 2006–2009, Including Breach Form	No. of Rev./Ed. Stimulus Passages Needed for Field Test in Spring 2005	No. of Rev./Ed. Stimulus Passages Needed for Op. 2006, Including Breach Form	No. of MC Stimulus-Based Rev./Ed. MC Items Needed for Field Test in Spring 2005	No. of MC Stimulus-Based Rev./Ed. MC Items Needed for Op. 2006, Including Breach Form	No. of Field Test Forms
5	30	16	*10	4–6	100	24	30
8	30	16	*10	4–6	100	24	30
Total	60	32	20	8–12	200	48	60

*Each of the 10 unique stimulus-based passages was split across two forms, with five unique multiple-choice items on each form per passage. This spread the total representation of passages and MC items across 10 forms (10 forms / 10 stimulus-based passages multiplied by 10 MC items for a total of 100 MC items). The 10 passages and 100 MC items were then repeated across the remaining 20 forms. Except for the field-test prompt portions of the test, Forms 01–10 were identical to forms 11–20 and forms 21–30.

The standalone field test design yielded enough stimulus-based multiple-choice items for use in building the grades 5 and 8 operational 2006 tests, including the breach test forms. The field test also yielded enough stimulus-based multiple-choice items to be used in public release item samplers and used as matrix items on future tests.

The standalone field test yielded enough writing prompts for the development of the spring 2006–2009 operational and breach tests. The field test also yielded enough prompts to be banked for use in the development of future public release item samplers.

GRADE 11 TEST (TRANSITION) PLAN: EMBEDDED FIELD TEST

The transition plan for grade 11 included the administration of the existing grade 11 writing test in the spring of 2005. This was the last administration of the existing grade 11 writing assessment. In addition, the spring 2005 operational grade 11 writing test included embedded field-test writing prompts and embedded stimulus-based multiple-choice items. Tables 5–2 and 5–3 provide an overview of the spring 2005 grade 11 writing test.

Table 5–2. 2005 Grade 11 Writing Operational Test Per Form, with Embedded Field-Test Stimulus-Based Multiple-Choice Items and Embedded 4 pt. Field-Test Writing Prompt

Grade	No. of Embedded FT Rev./Ed. MC per Op. Form	No. of FT Stimulus Passages per Op. Form	No. of Core 4-pt. Writing Prompts (WP) per Op. Form	No. of Embedded 4-pt. WP per Op. Form	Total No. of Items per Op. Form MC/WP	Total No. of Op. Forms
11	10	2	2	1	10/3	30

Table 5–3. 2005 Grade 11 Writing Operational Test Per 30 Forms

Grade	Total No. of Embedded FT Rev./Ed. MC per 30 Forms	Total No. of Unique Stimulus Passages per 30 Forms	Total No. of Core 4-pt. Writing Prompts (WP) per 30 Forms	Total No. of Embedded 4-pt. WP per 30 Forms	Total No. of MC and WP per 30 Forms MC/WP	Total No. of Op. Forms
11	100*	10*	2	30	100/32	30

*Each of the 10 unique stimulus-based passages was split across two forms, with five unique multiple-choice items on each form per passage. This spread the total representation of passages and MC items across 10 forms (10 forms / 10 stimulus-based passages multiplied by 10 MC items for a total of 100 MC items). The 10 passages and 100 MC items are then repeated across the remaining 20 forms. Except for the field-test prompt portions of the test, Forms 01–10 were identical to forms 11–20 and forms 21–30.

The total estimated administration time for the grade 11 operational test with embedded field test was approximately 190 minutes, broken out as 10 minutes for the multiple-choice field test items, 60 minutes for the field-test prompt, and 60 minutes for each of the two operational prompts.

The embedded field test yielded enough stimulus-based multiple-choice items for use in building the grade 11 operational 2006 tests, including the breach tests. The field test also yielded enough stimulus-based multiple-choice items to be used in public release item samplers and used as matrix items on future tests.

The embedded field test yielded enough writing prompts for the development of the spring 2006–2009 operational and breach tests. The field test also yielded enough prompts to be banked for use in the development of public release item samplers.

The spring 2005 grade 11 writing test was divided into three sessions, as follows:

- Session one included two embedded field-test stimulus-based passages, with five multiple-choice items per passage (ten total MC items), and one 4-pt. embedded field-test writing prompt. The MC stimulus-based passages were administered first.
- Session two included one 4-pt. operational writing prompt.
- Session three included one 4-pt. operational writing prompt.

Following the forms construction process, all forms were reviewed by PDE and revisions were made to items and/or format as needed. The forms were then printed, and packets of forms were spiraled and shipped according to the sampling plan described below. Each student participant received one form (a writing booklet). The writing booklet was used to collect demographic information, to record responses to the multiple-choice items, and to record responses to the writing prompts.

One administration manual for each grade was written and printed for district and school administrators and field test administrators.

ADMINISTRATOR QUESTIONNAIRE

A teacher (administrator) questionnaire was included in the packet of materials for the administrator. The purpose of the questionnaire was to gain further feedback on the items and directions for administration. In addition, those administering the field test were asked to estimate how long it took for students to complete the field test items. The completed administrator questionnaires were returned to DRC and the results were summarized.

SAMPLING

For all spring assessments, DRC Psychometric Services identified a district-level sampling plan that was used for field-test scoring. The sample was chosen randomly from a roster of all districts. This file contained descriptive information about the districts, including reported enrollment, number tested in 2004, mean PSSA scores for reading and math, percent non-white, and percent economically disadvantaged. The sampling was monitored to ensure it was representative of the state as a whole.

With the exception of Philadelphia and Pittsburgh, sampling was at district level, meaning all schools in a selected district were included in the sample. The two large districts were sampled at the school level, because it was not possible to draw a representative sample with those districts completely in or completely out of the sample.

For grade 11 writing, the field test was embedded in the regular operational administration with thirty forms spiraled across the state. Only districts or schools identified as included in the sample were scored on the field test writing samples. All embedded MC field test items were scored. The field test prompt was the first prompt (Prompt 1).

For grades 5 and 8 writing, the field test was conducted as a separate standalone administration. DRC's Performance Assessment Scoring (PAS) scored the sample districts (30 forms × 600 students = 18,000 documents to be scored per grade). Note that computer imaging was not used (as in the operational administration). Rather, intact test booklets were packeted and distributed for scoring with the scores entered into an electronic spreadsheet.

SAMPLE CHARACTERISTICS

Table 5–4 describes the characteristics of each sample by grade and content area. It also compares the sample achieved with the student population. When sampling is done at the district level and there is no record or penalty associated with participating in the field test, it is not possible to match the state demographics exactly. Overall, the sample size and participation were deemed sufficient to provide a defensible tryout of the items.

For a variety of reasons, not all schools and districts selected for the sample were able to participate in the field test. The table below describes the characteristics of the schools actually participating and returning data for the field tests. The data in the columns labeled *Operational* were taken from the 2005 Reading and Math Operational Assessment, which was conducted at a slightly different time. Because the grade 11 writing field test was embedded in the operational writing assessment, it matches the state data very closely. The grades 5 and 8 data were based on a district-level sampling plan implemented as a separate assessment and are not expected to match as closely. Also, the field test administrators were somewhat less conscientious about collecting and coding all the demographic fields than were the administrators responsible for the operational test. However, for the most part, the match between the field test characteristics and the state-wide operational characteristics was quite good.

Table 5–4. Match of Sample to State on Selected Demographics

		Grade 5		Grade 8		Grade 11	
	Group	Field Test	Opera-tional	Field Test	Opera-tional	Field Test	Opera-tional
Ethnic	White	80.8%	77.4%	82.4%	86.7%	80.0%	80.9%
	Black	10.6%	17.2%	10.9%	17.2%	12.8%	12.3%
	Hispanic	4.5%	6.5%	3.6%	6.2%	4.0%	3.9%
	Asian	1.9%	2.5%	1.7%	2.4%	2.4%	2.5%
	Native American	0.2%	0.1%	0.1%	0.1%	0.3%	0.1%
	Multi-Racial	2.1%	0.7%	1.3%	0.5%	0.3%	0.3%
IEP	No	85.2%	84.4%	86.3%	85.1%	87.0%	87.7%
	Yes	14.8%	15.6%	13.7%	14.9%	13.0%	12.3%
LEP	No	96.9%	96.9%	97.5%	97.8%	98.3%	98.3%
	Yes	3.1%	3.1%	2.5%	2.2%	1.7%	1.7%
504 Plan	No	99.2%	99.2%	99.3%	99.3%	99.4%	99.3%
	Yes	0.8%	0.8%	0.7%	0.7%	0.6%	0.7%
Economically Disadvantaged	No	65.4%	63.0%	68.7%	67.4%	77.1%	77.2%
	Yes	34.6%	37.0%	31.3%	32.6%	22.9%	22.8%

STATISTICAL ANALYSIS OF ITEM DATA

All field-tested items were analyzed statistically following conventional item analysis methods. For MC items the statistical indices known as traditional or classical item statistics included the point-biserial correlation (Pt Bis) for the correct and incorrect responses, percent correct (P-Value), and the percent responding to incorrect responses (distractors). For OE items the statistical indices included the item-total test correlation, the point-biserial correlation for each score level, percent in each score category or level, and the percent of non-scorable responses.

With any psychometric model, an item analysis is a search for unexpected results. In general, more capable students are expected to respond correctly to easy items and less capable students are expected to respond incorrectly to difficult items. If either of these situations does *not* occur, the item would be reviewed by DRC test development staff and committees of Pennsylvania educators to determine the nature of the problem and the characteristics of the students affected. The primary way of detecting such conditions is through the point-biserial correlation coefficient for dichotomous (MC) items and the item-total correlation for polytomous (OE) items. In each case, the statistic will be positive if the total test mean score is higher for the students who respond correctly to MC items (or attain a higher OE item score) and negative when the reverse is true.

Item statistics are used as a means of detecting items that deserve closer scrutiny, rather than being a mechanism for automatic retention or rejection. Toward this end, a set of criteria was used as a screening tool to identify items that needed a closer review by committees of Pennsylvania educators. For a MC item to be flagged, the criteria included any of the following:

- Point-biserial correlation for the correct response of less than 0.25
- Point-biserial correlation for any incorrect response greater than 0.0
- Percent correct less than 30% or greater than 90%
- Percent responding to any incorrect responses greater than the percent correct

For an OE item to be flagged, the criteria included any of the following:

- Item-total correlation less than 0.40
- Percent in any score category less than 10% or greater than 40%
- Non-scorable responses greater than 10 percent

Item analysis results for multiple-choice field test items are presented in Appendices E through G.

DIFFERENTIAL ITEM FUNCTIONING

Bias in test items can present itself in a variety of ways: through the language, the format, the content, or the behaviors required. It can result from membership in a specific subpopulation or from factors correlated with a subpopulation. It can affect all members of the subpopulation, or it can affect only those in specific ranges of ability. Understanding how bias arises and how it presents itself has an effect on how best to detect and correct it.

LIMITATIONS OF STATISTICAL DETECTION

No statistical procedure should be used as a substitute for rigorous, hands-on reviews by content and bias specialists. The statistical results can help to organize the review so the effort is concentrated on the most problematic cases; however, no items should be automatically rejected simply because a statistical method flagged them or accepted because they were not flagged.

Statistical detection of item bias is an inexact science. There have been a variety of methods proposed for detecting bias, but no one statistic can be considered either necessary or sufficient. Different methods are more or less successful depending on the situation. No analysis can guarantee that a test is free of bias, but almost any thoughtful analysis will uncover the most flagrant problems.

A fundamental shortcoming of all of the statistical methods used to detect DIF is that all are intrinsic to the test being evaluated. If a test is unbiased overall but contains one or two biased items, any method will locate the problems. If, however, all items on the test are consistently biased against a subpopulation, a statistical analysis of the items will not be able to separate bias effects from true differences in achievement.

MANTEL-HAENSZEL PROCEDURE FOR DIFFERENTIAL ITEM FUNCTIONING

The *Mantel-Haenszel* procedure for detecting differential item functioning is the most commonly used technique in educational testing. It does not depend on the application or the fit of any specific measurement model. However, it does have significant philosophical overlap with the Rasch model since it uses total score to organize the analysis.

Differential item functioning is present when examinees of equal ability but different subgroup membership do not have the same chance of answering the item correctly. If this inequity is associated with gender or ethnic groups, the item could be described as potentially biased.

The procedure as implemented by DRC contrasts a focal group with a reference group. While it makes no practical difference in the analysis which group is defined as the focal group, the protected class or the group most apt to be disadvantaged by a biased measurement is typically defined as the focal group. The Mantel-Haenszel (MH) statistic (Mantel & Haenszel, 1959) for each item is computed from a contingency table. It has two groups (focal and reference), two outcomes (right or wrong), and ability groupings. The ability groups are defined by the score distribution for the total examinee populations.

The basic MH statistic is a single degree of freedom chi-square that compares the observed number in each cell to the expected number. The expected counts are computed to ensure that the analysis is not confounded with differences in the achievement level of the two groups.

For constructed response items, a comparable statistic is computed based on the standardized mean difference (SMD) (Dorans, Schmitt & Bleistein, 1992), computed as the differences in mean scores for the focal and reference groups if both groups had the same score distribution.

To assist the review committees in interpreting the analyses, the items are assigned a severity code based on the magnitude of the MH statistic. Items classified as A+ or A- have no statistical indication of differential item functioning. Items classified as B+ or B- have some indication of DIF and may not require revision. Items classified as C+ or C- have strong evidence of DIF and should be reviewed and revised if they are to be used again. The plus sign indicates that the item

favors the focal group and a minus sign indicates that the item favors the reference group.

Counts of the number of items from each grade that were assigned to each severity code are shown below in Table 5-5.

Table 5-5. 2006 DIF Summary

2006 Multiple Choice Item DIF Summary

Multiple Choice Item Male/Female DIF Counts						
	A	**B-**	**B+**	**C-**	**C+**	**Total**
5	64	0	0	0	0	64
8	63	0	1	0	0	64
11	63	0	0	0	1	64

Multiple Choice Item White non-Hispanic—Black/African American non-Hispanic DIF Counts						
	A	**B-**	**B+**	**C-**	**C+**	**Total**
5	51	12	0	1	0	64
8	59	4	0	1	0	64
11	54	7	0	3	0	64

Multiple Choice Item White non-Hispanic—Latino/Hispanic DIF Counts						
	A	**B-**	**B+**	**C-**	**C+**	**Total**
5	50	9	0	5	0	64
8	58	4	1	1	0	64
11	53	7	1	3	0	64

Multiple Choice Item White non-Hispanic—Asian or Pacific Islander DIF Counts						
	A	**B-**	**B+**	**C-**	**C+**	**Total**
5	51	1	7	3	2	64
8	55	0	4	1	4	64
11	50	4	3	4	3	64

Note: Writing prompts were field tested and analyzed for DIF in spring of 2005. See the Technical Report from that year for additional information about these items.

REVIEW OF ITEMS WITH DATA

In the preceding section on Statistical Analysis of Item Data, it was stated that test development content-area specialists used certain statistics from item and DIF analyses of the 2005 field test to identify items for further review. Specific flagging criteria for this purpose were specified in the section above. Items not identified for this review were those that had good statistical characteristics and, consequently, regarded as statistically acceptable. Likewise, items of extremely poor statistical quality were easily regarded as unacceptable and needed no further review. However, there were some items, relatively few in number, which DRC content-area test development specialists deemed as needing further review by a committee of Pennsylvania educators. The intent was to capture all items that needed a closer look; thus the criteria employed tended to over-identify rather than under-identify items.

The review of the items with data was conducted by subject-area content committees composed of 14 Pennsylvania teachers, administrators, and PDE staff. The review took place on July 11, 2005. In this session committee members were first trained by Dr. Ronald Mead, DRC Senior Psychometrician, with regard to the statistical indices used in item evaluation. This was followed by a discussion with examples concerning reasons that an item might be retained regardless of the statistics. The committee review process involved a brief exploration of possible reasons for the statistical profile of an item (such as possible bias, grade appropriateness, instructional issues, etc.) and a decision regarding acceptance. DRC content-area test development specialists facilitated the review of the items.

Chapter Six: Operational Forms Construction for 2006

FINAL SELECTION OF ITEMS AND 2006 PSSA FORMS CONSTRUCTION

When the final selection of items for the operational 2006 test was ready to begin, the candidate items that emerged from the Spring 2005 field test had undergone multiple reviews, including:

- Reviews by DRC content-area test development specialists and curriculum specialists
- Formal bias, fairness, and sensitivity review by the Bias, Fairness, and Sensitivity Committee consisting of an expert, multi-ethnic group of men and women with members also having expertise with special needs students and English Language Learners
- Formal review by the content committees consisting of Pennsylvania educators, including teachers as well as district personnel
- PDE review
- Item data review by members of the PDE subject-area teacher committees

The end product of the above process was an "item status" designation for each field-tested item. All items having an item status code of "Acceptable" were candidates to be selected for the 2006 PSSA. To have an item status code of "Acceptable" meant that the item met the following criteria:

- Appropriately aligned with its Academic Standard
- Acceptable in terms of bias/fairness/sensitivity issues, including differential item functioning (for gender and race)
- Free of major psychometric flaws, including a special review of flagged items

Next, all relevant information regarding the acceptable items was entered into the IVAN system. From the IVAN system, Excel files were created for writing. These files contained all relevant content codes and statistical characteristics. The IVAN system also created for each acceptable item a card displaying the item and all relevant content codes and item statistics for use by the content-area test development specialists and psychometric services staff.

DRC test development specialists reviewed the test design blueprint, including the number of items per content standard. Considerations such as item focus, topic variety, and answer option distribution were also noted.

Psychometricians provided content-area test development specialists with an overview of the psychometric guidelines for forms construction, including guidelines for selecting linking items to link to previous test forms.

Senior DRC content-area test development specialists reviewed all items in the operational pool to make an initial selection for common (core) items (including writing prompts) and matrix items and passages according to test blueprint requirements and psychometric guidelines. No changes were made to any item since even slight alterations could affect how an item performs on subsequent testing.

For the common items, this meant that the combination of MC items and writing prompts would

tap an appropriate variety of components under the Academic Standards under each reporting category. Items were selected in sets combined under the umbrella of a stimulus passage, and were examined with regard to how well they went together as a set. Of particular concern were the following:

- One item providing cues as to the correct answer to another item
- Presence of "clang" (distractors not unique from one another)
- Diversity of names and topics for gender and ethnicity

The first round of items was then evaluated for statistical features such as an acceptable point-biserial correlation and whether the items, as a collection, had a correct answer distribution of approximately 25 percent in each of the four positions. Selected items that were psychometrically problematic resulted in a search by the senior reviewer for suitable replacements. At this point, the second round of items was analyzed. If necessary, this iterative process between content-based selections and statistical properties continued in an effort to reach the best possible balance.

The process for selecting operational matrix (linking) MC items was a little different. The chief consideration was that items in the matrix section of the various forms, together with the common items, would yield a greater overall pool of items from which reliable results could be generated for school-level reporting. Once again the cardinal principle was the selection of an appropriate number of items to properly cover the Academic Standards. The test development specialist's task was to distribute these items so that items assigned to a particular form would go well with one another and reflect the same content and statistical considerations as previously outlined. Additionally, the forms needed to display similar difficulty levels.

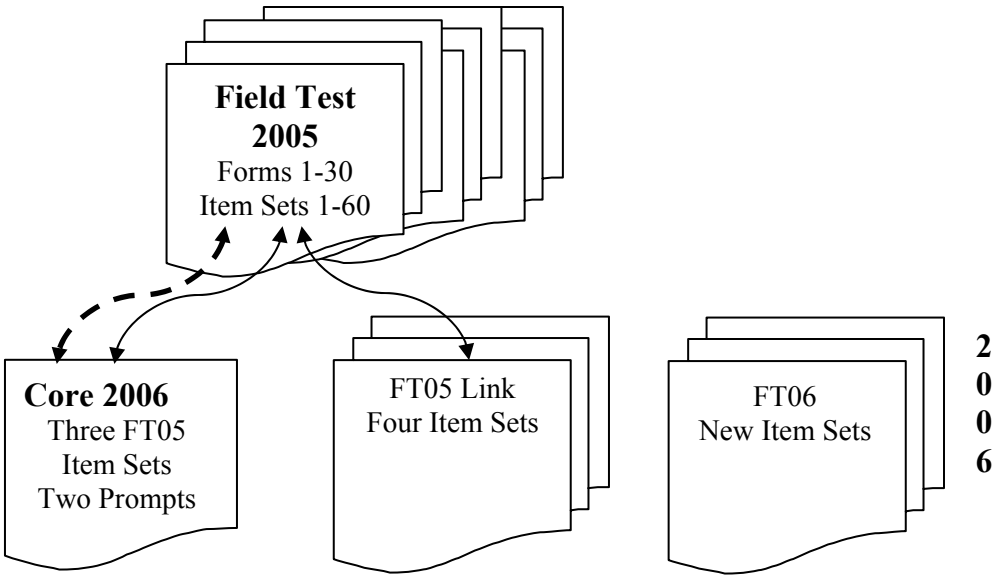
Once the recommendations were finalized for the common/core and matrix items, they were submitted to PDE for review. Department staff provided feedback, which could be in the form of approval or recommendations for replacing certain items. Any item replacement was accomplished by the collective effort of the test development specialists, psychometricians, and PDE staff until final PDE approval.

THE LINKING PROCESS

Year-to-year linking (equating) is accomplished primarily with multiple-choice items moving from field test to matrix. Multiple-choice matrix items are maintained in the same location and the same context as they were used at field test. An alternate route for linking is established through the use of the writing prompts which were field tested in 2005. There are 4 unique designated multiple-choice matrix items per grade appearing on 4 out of the 10 forms for a total of 16 unique matrix items per grade. There are 2 core prompts available for linking per grade.

The following 2 tables illustrate how the linking items connect the assessment across administrations.

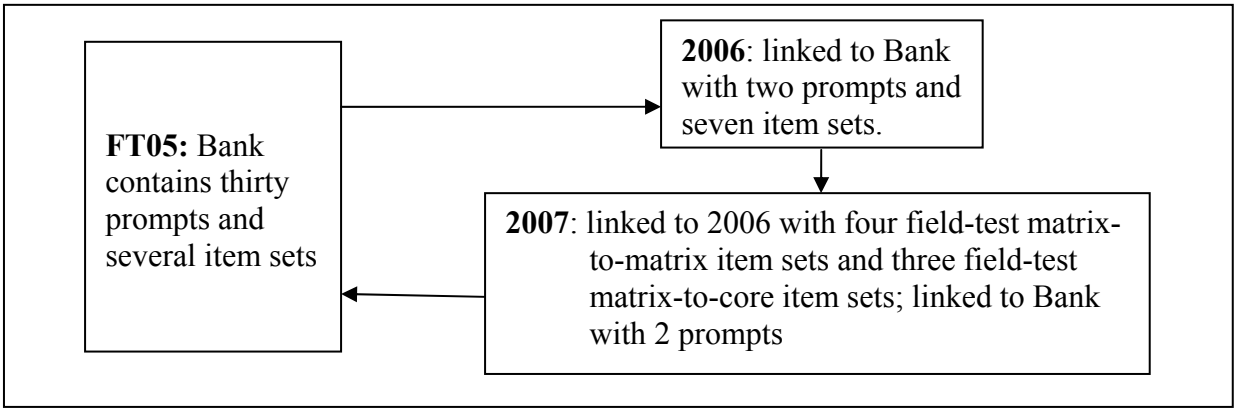
Table 6–1. Linking Plan for PSSA Writing from 2005 to 2006



Notes:

1. Solid-line connectors represent links of item sets, each including a passage of four passage-based multiple choice items.
2. Dash-line connectors represent links containing two writing prompts.
3. Horizontal rows represent administrations (years).
4. All years will be linked to FT05 through the prompts, providing an alternate path for year-to-year linking.
5. This design incorporates consistency checks: *Bank→2006→2007→Bank* should sum to zero, statistically. Similarly, for 2007 to 2008 and 2008 to 2009.

Table 6–2. Linking Plan for PSSA Writing for 2005 through 2007



EMBEDDED FIELD-TEST ITEMS

The 2006 PSSA test forms contained common items (identical on all forms) along with matrix/embedded field-test items. The common items are a set of "core" items taken by all students. The matrix and field-test items are embedded and are unique, in most instances, to a form; however, there are several instances in which a matrix or embedded field-test MC item appears on more than one form. There were no open-ended field test items in this administration; both prompts were common across forms. The purpose of administering field test items is to obtain statistics for new items which are then reviewed before becoming operational.

SPECIAL FORMS USED IN THE 2006 PSSA

BRAILLE AND LARGE PRINT

Students with visual impairments were able to respond to test materials that were available in either **Braille** or **large print**. At each grade level assessed, one form was selected for the creation of a Braille and a large-print edition. School district personnel ordered Braille or large-print assessment materials directly from the Pennsylvania Training and Technical Assistance Network (PaTTAN) in Harrisburg. They could also contact PaTTAN for technical assistance regarding students with visual impairments.

School personnel were directed to transcribe all student answers (MC and WP) into scannable answer documents exactly as the student responded. No alterations or corrections of student work were permitted, and the answer document had to have the identical form designation.

Instructions for the appropriate use of these special forms are detailed in the *2006 Accommodations Guidelines* (PDE, December 2005) available on the PDE website at www.pde.state.us.

Chapter Seven: Test Administration Procedures

TEST SESSIONS, TIMING, AND LAYOUT

The test window for the 2006 operational assessment was from February 13 through February 24, 2006, including make-ups. The assessment consisted of three sections. Additional information concerning testing time and test layouts can be found in Chapter 3.

SHIPPING, PACKAGING, AND DELIVERY OF MATERIALS

There were two shipments sent out by Data Recognition Corporation (DRC). Shipment one was delivered by January 13, 2006, and contained the *Handbook for Assessment Coordinators and Administrators* and the *Directions for Administration* for each grade tested at a school. Shipment two was delivered by January 30, 2006, and contained the administrative materials (e.g., return shipping labels and student precode labels) and secure materials (e.g., writing booklets). DRC ensured that all assessment materials were assembled correctly prior to shipping. DRC Operations staff used the automated Operations Materials Management System (Ops MMS) to assign secure materials to a district at the time of ship out. This system used barcode technology to provide an automated quality check between items requested for a site and items shipped to a site. A shipment box manifest was produced for and placed in each box shipped. DRC Operations staff double checked all box contents with the box manifest prior to the box being sealed for shipment to ensure accurate delivery of materials. DRC Operations staff performed lot acceptance sampling on both shipments. Districts and schools were selected at random and examined for correct and complete packaging and labeling. This sampling represented a minimum of 10 percent of all shipping sites.

DRC used UPS to ship materials to districts. DRC's materials management system, along with the shipper's system, allowed DRC to track the items from the point of shipment from DRC's warehouse facility to receipt at the district, school, or testing site. All DRC shipping facilities, materials processing facilities, and storage facilities are secure. Access is restricted by security code. Non-DRC personnel are escorted by a DRC employee at all times. Only DRC inventory control personnel have access to stored secure materials. DRC employees are trained and made aware of the high level of security that is required.

DRC packed more than 509,682 writing booklets, 61,551 manuals, and 71,592 non-secure materials for over 2,938 schools. DRC used UPS to deliver 8,596 boxes of materials to the testing sites.

MATERIALS RETURN

The materials return window was February 27, 2006 through March 20, 2006. DRC used UPS for all returns.

TEST SECURITY MEASURES

Test security is essential to obtaining reliable and valid scores for accountability purposes. The 2006 PSSA included a Test Security Affidavit that was to be signed and returned by every principal or director where testing materials were shipped. 2,871 of the Test Security Affidavits for the Writing assessment that were sent to a total of 3,201 testing sites were signed and returned to DRC. The purpose of the affidavit was to serve as a tool to document that the

individuals responsible for administering the assessments both understood and acknowledged the importance of test security and accountability. The affidavit attested that all security measures were followed concerning the handling of secure materials. Some of the security measures included:

1. The contents of the test were not discussed, disseminated, described, or otherwise revealed to anyone.
2. The contents of the test were not kept, copied, or reproduced.
3. All booklets were kept in a locked, secure storage area at both the district and school levels.

SAMPLE MANUALS

Copies of the *Handbook for Assessment Coordinators and Administrators* and the *Directions for Administration* can be found on the Pennsylvania Department of Education website at www.pde.state.pa.us.

ASSESSMENT ACCOMMODATIONS

An accommodations manual entitled *2006 Accommodations Guidelines* (PDE, December 2005) was developed for use with the 2006 PSSA. Additional information regarding assessment accommodations can be found in Chapter 4.

Chapter Eight: Processing and Scoring

RECEIPT OF MATERIALS

Receipt of PSSA Writing materials began on February 27, 2006, and concluded on March 20, 2006. DRC's Operations Material Management System (Ops MMS) was utilized to receive secure materials securely, accurately, and efficiently. This system features advanced automation and cutting-edge barcode scanners. Captured data were organized into reports, which provided timely information with respect to suspected missing material.

The first step in the Ops MMS was the Box Receipt System. When a shipment arrived at DRC, the boxes were removed from the carrier's truck and passed under a barcode reader, which read the barcode contained on the return label and identified the district and school. If the label could not be read automatically, a floor operator entered the information into the system manually. The data collected in this process were stored in the Ops MMS database. After the barcode data were captured, the boxes were placed on a pallet and assigned a corresponding pallet number. A "three way match" among the district box count, the carrier box count, and the DRC return box count was conducted to verify a box return accuracy rate of 100%.

Once the box receipt process was completed, the materials separation phase began. Warehouse personnel opened the district boxes and sorted the contents by grade and status (used/unused) into new boxes. Once filled, a sorted box's documents were loaded into an automated counter, which recorded a booklet count for each box. An on-demand DRC box label was produced that contained a description of each box's contents and quantity in both barcode and human-readable format. This count remained correlated to the box as an essential quality control step throughout secure booklet processing and provided a target number for all steps of the check-in process.

Once labeled, the sorted and counted boxes proceeded to booklet check-in. This system used streamfeeder automation to carry documents past oscillating scanners that captured data from up to two representative barcodes and stored it in the Ops MMS database.

- The secure booklet check-in operator used a hand scanner to scan the counted box label. This procedure input material type and quantity parameters for what the Ops MMS should expect within a box. It then loaded the box's contents into the streamfeeder.
- The documents were fed past oscillating scanners that captured either a security code or both a security code and a pre-code, depending upon material type. A human operator monitored an Ops MMS screen, which displayed scan errors, an ordered accounting of what was successfully scanned, and the document count for each box.
- When all materials were scanned and the correct document count was reached, the box was sealed and placed on a pallet. If the correct document count was not reached, or if the operator encountered difficulties with material scanning, the box and its contents were delivered to an exception handling station for resolution.

This check-in process occurred immediately upon receipt of materials; therefore, DRC provided immediate feedback to districts and schools regarding any missing materials based on actual receipts versus expected receipts.

Upon completion of secure booklet check-in, DRC produced a Missing Materials Report that listed all schools in each participating district and any booklets not returned to DRC listed by school and security number.

After scannable materials were processed through Book Receipt, the materials became available to the DRC Document Processing Center Log-In staff for document log-in. Based on a pre-determined sampling and calibration plan, the staff prioritized answer documents using the following process:

- A DRC scannable barcode batch header was scanned, and a batch number was assigned to each box of answer documents.
- The DRC box label barcode was scanned into the system to link the box and writing documents to the newly created batch and to create a Batch Control Sheet.
- The DRC box label barcode number, along with the number of writing documents in the box, was printed on the Batch Control Sheet for document tracking purposes. All documents that were linked to the box barcode were assigned to the batch number and tracked through all processing steps. As documents were processed, DRC staff dated and initialed the Batch Control Sheet to indicate that proper processing and controls were observed.
- Before the answer documents were scanned, all batches went through a quality inspection to ensure batch integrity and correct document placement.

After a quality check in the DRC Document Processing Log-in area, the spines were cut off the scannable documents, and the pages were sent to DRC's Imaging and Scoring System.

SCANNING OF MATERIALS

DRC used its image scanning system to capture constructed-response items as images. These were then loaded into the image scoring system for both the handscoring of constructed-response items and for the capture of multiple-choice and demographic data.

DRC's image scanners were calibrated using a standard deck of scannable pages with 16 known levels of gray. On a predefined page location, the average pixel darkness was compared to the standard calibration to determine the level of gray. Marks with an average darkness level of 4 or above on a scale of 16 (0 through F) were determined to be valid responses, per industry standard. If multiple marks were read for a single item and the difference of the grayscale reads was greater than four levels, the lighter mark was discarded. If the multiple marks had fewer than four levels of grayscale difference, the response was flagged systematically and forwarded to an editor for resolution.

Customized scanning programs for all scannable documents were prepared to read the writing documents and to electronically format the scanned information. Before materials arrived, all image scanning programs went through a quality review process that included scanning of mock data from production booklets to ensure proper data collection.

- DRC's image scanners read selected-response, demographic, and identification information. The image scanners also used barcode readers to read pre-printed barcodes from a label on the booklet.
- The scannable documents were automatically fed into the image scanners where pre-defined processing criteria determined which fields were to be captured electronically. Constructed-response images were separated out for image-based scoring.
- During scanning, a unique serial number was printed on each sheet of paper. This serial

number was used for document integrity and to maintain sequencing within a batch of writing documents.

- A monitor randomly displayed images, and the human operator adjusted or cleaned the scanner when the scanned image did not meet DRC's strict quality standards for image clarity.
- All images passed through a process and a software clean-up program that despeckled, deskewed, and desmeared the images. A random sample of images was reviewed for image quality approval. If any document failed to meet image quality standards, the document was returned for rescanning.
- Page scan verification was performed to ensure that all pre-defined portions of the answer documents were represented in their entirety in the image files. If a page was missing, the entire writing document was flagged for resolution.

After each batch was scanned, writing documents were processed through a computer-based edit program to detect potential errors as a result of smudges, multiple marks, and omits in predetermined fields. Marks that did not meet the pre-defined editing standards were routed to editors for resolution.

- Experienced DRC Document Processing Center Editing staff reviewed all potential errors detected during scanning and made necessary corrections to the data file. The imaging system displayed each suspected error. The editing staff then inspected the image and made any needed corrections using the unique serial number printed on the document during scanning.
- Upon completion of editing, quality control reports were run to ensure that all detected potential errors were reviewed again and a final disposition was determined.

Before batches of writing documents were extracted for scoring, a final edit was performed to ensure that all requirements for final processing were met. If a batch contained errors, it was flagged for further review before being extracted for scoring and reporting.

- During this processing step, the actual number of documents scanned was compared to the number of writing documents assigned to the box during book receipt. Count discrepancies between book receipt and writing documents scanned were resolved at this time.
- Once all requirements for final processing were met, the batch was released for scoring and student level processing.

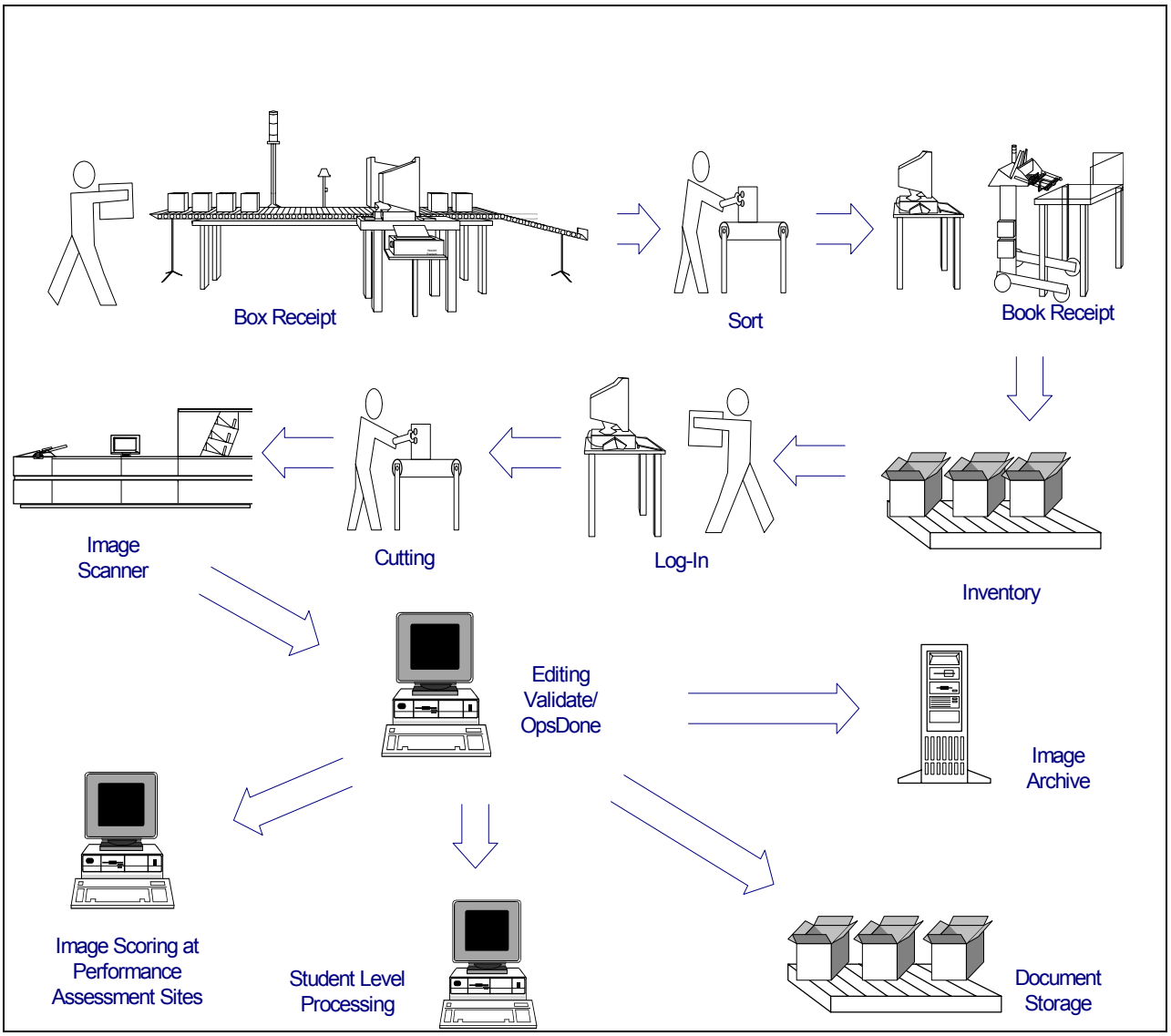
Table 8-1 shows the number of writing booklets received through booklet check in and the number of booklets that contained student responses that were scanned and scored.

Table 8-1. Counts of 2006 PSSA Writing Materials Received – Grades 5, 8, and 11

	Writing Booklets Received	Used Writing Booklets Scanned
Grade 5	166,613	135,119
Grade 8	174,262	148,300
Grade 11	168,736	139,760

Figure 8.1 illustrates the production workflow for DRC's Ops MMS and Image Scanning and Scoring System from receipt of materials through all processing of materials and the presentation of scanned images for scoring.

Figure 8.1. Workflow System



MATERIALS STORAGE

Upon completion of processing, student answer materials are boxed for security purposes and final storage:

- Project-specific box labels were created containing unique customer and project information, materials type, batch number, pallet/box number, and the number of boxes for a given batch.
- Boxes were stacked on project-specific pallets that were labeled with a list of its contents and delivered to the Materials Distribution Center for final secure storage.
- Materials will be destroyed one year after contract year ends with PDE written approval.

SCORING MULTIPLE-CHOICE ITEMS

The scoring process included the scoring of multiple-choice items against the answer key and the aggregation of raw scores from the constructed responses. A student's raw score is the actual number of points achieved by the student for tested elements of an assessment. From the raw scores, the scale scores were calculated.

The student file was scored against the finalized and approved multiple-choice answer key. Items were scored as right, wrong, omitted, or double-gridded (more than one answer was bubbled for an item). Sections of the test were evaluated as a whole and an attempt status was determined for each student. The score program defined all data elements at the student level for reporting.

RANGEFINDING

Once student answer documents were received and processed, DRC's Performance Assessment Services (PAS) staff began to assemble groups of responses that exemplified the different score points represented in the mode-specific and conventions scoring guidelines. This was done for both the common persuasive and informational prompts in grades 11 and 8 and for the narrative and informational prompts in grade 5. 2006 was the first year these new scoring guidelines were used for all three grades of the operational assessment.

Once examples for all the score points were identified, sets of items were put together by mode. These sets included all of the 2005 field test prompts that were selected for the 2006 operational test. These sets were copied for use at rangefinding, held March 15–16, 2006 at the Holiday Inn, Grantville, Pennsylvania. The rangefinding committees consisted of Pennsylvania educators, PDE staff members, DRC Test Development staff, and DRC Performance Assessment Services.

After an introductory general session, committees broke into grade level groups. Copies of the student example sets were presented to the committees by mode. The committees reviewed and scored the student samples together to ensure that everyone was interpreting the scoring guidelines consistently. Committee members then went on to score responses independently and those scores were discussed until a consensus was reached. Only responses for which a good agreement rate was attained were used in training the readers. Discussions of the responses used the language of the scoring guidelines, assuring PDE and all involved that the score point examples clearly illustrated the specific requirements of each score level. DRC PAS staff made notes of how and why the committees arrived at score point decisions, and this information was used by the individual scoring directors in reader training.

READER RECRUITMENT/QUALIFICATIONS

DRC retains a number of experienced readers from year to year, and those readers made up approximately 60% of the reader pool (N=250) for 2006. To complete the reader staff for this project, DRC placed advertisements in local papers, minority publications, teacher newsletters, regional colleges, and universities. Open houses were held and applications for reader positions were screened by the DRC recruiting staff. Candidates were personally interviewed and a mandatory, on-demand writing sample was collected, along with references and proof of a four-year college degree. In this screening process, preference was given to candidates with previous experience scoring large-scale assessments and with degrees emphasizing expertise in writing. Since readers had to have a strong content-specific background, the reader pool consisted of educators, writers, editors, and other professionals who were valued for their experience, but who were also required to set aside their own biases about student performance and accept the

scoring standards. All readers on this assessment held at least a four-year degree.

LEADERSHIP RECRUITMENT/QUALIFICATIONS

Scoring directors and team leaders were chosen by the project director from a pool consisting of experienced individuals who had proved to be successful readers and leaders on previous DRC contracts. Selectees had strong backgrounds in both scoring and writing and demonstrated organization, leadership, and management skills. The scoring directors and a majority of the team leaders had at least five years of leadership experience on the PSSA. All scoring directors, team leaders, and readers were required to sign confidentiality forms before any training or handling of secure materials began.

Each room of readers was assigned a scoring director. This individual was monitored by the project content coordinator and led the hand scoring for the duration of the project. The scoring director assisted in rangefinding, worked with supervisors to create training materials, conducted the team leader training, and was responsible for training the readers. The scoring director also made sure that reports were available and interpreted reports for the readers. The scoring director supervised the team leaders.

Team leaders assisted the scoring director with reader training and monitoring by working with their teams in small group discussions and answering individual questions that readers may not have felt comfortable asking in a large group. Once readers had qualified, the team leaders were responsible for maintaining the accuracy and workload of team members. The ongoing monitoring identified those readers who were having difficulty with scoring accurately and resulted in the reader receiving one-on-one retraining or in pairing that reader with a stronger reader. This process corrected any inaccuracies in scoring and, if not, that reader was released from the project.

TRAINING

After rangefinding was completed, DRC's PAS staff compiled the scoring guidelines and the scored student examples from the committees into packets used for training the readers. Responses that were relevant in terms of the scoring concepts they illustrated were annotated for use in a scoring guide. The scoring guide for each mode served as the reader's constant reference. Readers were instructed how to apply the guidelines and were required to demonstrate a clear comprehension of each Assessment Anchor set by performing well on the training materials that were presented for each grade and mode. Training and qualifying sets consisted entirely of examples of student responses chosen by the rangefinding committee.

Team leaders assisted the scoring directors with the training and monitoring of readers. The scoring director conducted the team leader training before the reader training. This training followed the same procedures as the reader training, but qualifying standards were more stringent because of the responsibilities required of the team leaders. During their training, all materials were reviewed and discussed and anticipated reader questions and concerns were addressed. Team leaders were required to annotate all of their training responses with the official annotations received from the content committee members at the rangefinding meetings. To facilitate scoring consistency, it was imperative that each team leader imparted the same rationale for each response that other team leaders used. Once the team leaders qualified, leadership responsibilities were reviewed and team assignments were given. A ratio of one team leader for each group of 8–10 readers ensured adequate monitoring of the readers.

Reader training began with the scoring director providing an intensive review of the scoring guides and anchor papers to all readers. Next, the readers "practiced" by independently scoring the responses in the training sets. Afterwards, the scoring director and team leaders led a thorough discussion of each set, in either a small group or room-wide setting.

Once the scoring guides and all the training sets were discussed, readers were required to apply the scoring criteria by qualifying (i.e., scoring with acceptable agreement to the "true" scores) on at least one of the qualifying sets. Readers who failed to achieve the level of agreement determined by PDE were given additional training to acquire the highest degree of accuracy possible. Readers who did not perform at the required level of agreement by the end of the qualifying process were not allowed to score "live" student work and were released from the project.

HANDSCORING PROCESS

Student responses were scored independently and by multiple readers. All essays, grades 5, 8, and 11, were read once with 10% scored twice to ensure reliability. PDE determined the required number of reads.

Readers scored the imaged student responses on PC monitors at the DRC Scoring Center in Woodbury, Minnesota. Readers were seated at tables with two imaging stations at each table. Image distribution was controlled, thus ensuring that they were sent to designated groups of readers qualified to score those prompts. Imaged student responses were electronically separated for routing to individual readers by prompt, and readers were only provided with student responses for which they were qualified to score. Readers read each response and keyed in the scores. Alerts and non-score mismatches were routed to the scoring director or team leaders for electronic review and resolution.

Upon completion of operational scoring, 15,855 grade 11 students fell within the 15 point rescore indicator. The essays of these students were each read again with 10% scored twice to ensure reliability. PDE determined the point rescore indicator and required number of reads.

QUALITY CONTROL

Reader accuracy was monitored throughout the scoring session by producing both daily and on-demand reports, ensuring that an acceptable level of scoring accuracy was maintained. Inter-reader reliability was tracked and monitored with multiple quality control reports that were reviewed by quality assurance analysts. These reports were generated at the handscoring center and were reviewed by the scoring directors, team leaders, project coordinators, and project directors. The following reports were used in scoring the 2006 writing portion of the PSSA:

- **The Reader Monitor Report** monitored how often readers were in exact agreement and ensured that an acceptable agreement rate was maintained. This report provided daily and cumulative exact and adjacent inter-reader agreement and the percentage of responses requiring resolution. (see Table 8-2)
- **The Score Point Distribution Report** monitored the percentage of responses given each of the score points. For example, this daily and cumulative report showed how many 0s, 1s, 2s, 3s, and 4s a reader had given to all the responses he or she had scored at the time the report was produced. It also indicated the number of responses read by each reader so that production rates could be monitored.

- **The Item Status Report** monitored the progress of handscoring. This report tracked each response and indicated the status (e.g., "needs second reading," "complete"). This report ensured that all discrepancies were resolved by the end of the project.
- **The Response Read by Reader Report** identified all responses scored by an individual reader. This report was useful if any responses needed rescoring because of reader drift.
- **The Read-Behind Log** was used by the team leader/scoring director to monitor reader reliability. Student responses were randomly selected and team leaders read scored items from each team member. If the team leader disagreed with the reader's score, remediation occurred. This proved to be a very effective type of feedback because it was done with "live" items scored by a particular reader.

Recalibration sets were used throughout the scoring sessions to monitor scoring by comparing each reader's scores with the true scores and to refocus readers on Pennsylvania scoring standards. This check made sure there was no change in the scoring pattern as the project progressed. Readers failing to achieve a certain percent of agreement with the recalibration true scores were given additional training to achieve the highest degree of accuracy possible. Readers who were unable to recalibrate were released from the project. The procedure for creating and reading recalibration sets was similar to the one used for the training sets.

To handle possible alert papers (i.e., student responses indicating potential issues related to the student's safety and well-being that may require attention at the state or local level), the imaging system allowed readers to forward responses needing attention to the scoring director. These alerts were reviewed by the project director, who then notified PDE of this occurrence. However, PDE did not receive the student's responses or any other identifying information on that student. Also, at no time did the reader know anything about the student's identity.

Table 8-2 shows the exact and adjacent agreement rates of the readers for the writing essays in grades 5, 8, and 11.

Table 8–2 Inter-rater Agreement for 2006 Grades 5, 8, & 11 Writing

		Composition [score range of 1-4]			Revising and Editing [score range of 1-4]		
Grade	Prompt	% Exact Agreement	% Adjacent Agreement	% Exact + Adjacent	% Exact Agreement	% Adjacent Agreement	% Exact + Adjacent
5	1	75	25	100	73	27	100
	2	75	25	100	74	26	100
8	1	76	24	100	73	27	100
	2	77	23	100	72	28	100
11	1	77	23	100	72	28	100
	2	78	22	100	74	26	100

MATCH-BACK RULES

In order to create a single student record in the central student file, it was necessary to establish match-back rules to combine separate student records into one student record. Match-back rules were applied to link multiple-choice and constructed responses. They were also used to merge student responses captured on different subjects and to link test results with student demographic information.

DATA EXCHANGE, STORAGE, AND RECOVERY POLICIES

DATA EXCHANGE PROCEDURES

The exchange of data between DRC, PDE, and other contractors is a critical and essential component in the success of the PSSA program. To support this process, DRC used the following data exchange procedures to ensure that all data files were successfully and accurately transferred.

- Files were posted to DRC's secure Pennsylvania FTP site with a standard and logical folder structure.
- Standard file naming conventions were established and used.
- The information necessary to perform these quality control procedures accompanied each data exchange.

Data Exchange Quality Control Procedures

- ***Record Count Check –*** Confirm the expected record count and provide the record count in files sent and received.
- ***File Count Check–*** Confirm that the number of files sent and received matches the number of files expected.
- ***Duplicate File Check –*** Verify that duplicate files were not sent or received.
- ***File Date*** – Verify that the version of the file received matches the file creation date.
- ***File Type Verification Check –*** Verify that data sent and received matches the format expected (e.g., Excel, CSV, PDF, Text file [delimited/fixed field length]).
- ***File Log –*** A log of files sent and received will be maintained.
- ***Data Validation –*** Data checking procedures will be used to verify that the data is in the specified file layout and matches the expected values.

IMAGES

As part of the scanning process, the multi-page TIFF images were archived to tape before being separated into single page TIFFs and transmitted to the scoring centers. If any of the images were lost/deleted/corrupted at a scoring center, they could be restored from the archived multi-page TIFF images. In addition to archiving the images, the scoring center servers used RAID (Redundant Array of Independent Disks) 5 disk management technology to mirror the images to redundant disk drives. If a disk drive failed in a scoring center server, the images could be quickly restored from the redundant disk drive. In the event that the disk drive and the multi-page TIFF images could not be restored, the original documents would be rescanned. Images are stored for a PDE specified period.

DATA

Once a reader submitted a score for a constructed-response item, the data was electronically transmitted to our SQL Servers. The log files documenting the changes were backed up hourly. Full back-ups were done nightly (Monday–Friday) and two additional full back-ups were run

over the weekend on the handscoring SQL Servers with the backup tapes being rotated off-site. All data is stored for a PDE-specified period.

STORAGE

All physical servers are housed in secure server rooms in DRC's corporate headquarters in Maple Grove, or the Brooklyn Park or Woodbury locations. The server rooms are constructed of concrete floors, walls, and ceilings and designed to be fire and crush proof. They have fire suppression systems to minimize the effect of any fire started within the server room. Access to the server rooms is controlled through a card access system and is restricted to authorized technology support staff only. A log is maintained documenting each time a server room is entered, by whom, and for what purpose. In case of a disaster at any of the locations, another server can take over full operations.

DRC maintains backup servers that can be used to replace a failed server within 24 hours. Every server's configuration is documented in the event a rebuild is required. Each server has an assigned primary and secondary network analyst responsible for its operation.

The servers utilize load-sharing, redundant power supplies and implement RAID subsystems to minimize the effect of a failed disk. The server rooms all have Uninterruptible Power Supply (UPS) systems. For longer periods of power failure, an on-site diesel power generate will automatically start and supply needed power. The computing environment, both servers and communications hardware, will continue to function without interruption when the utility power is disrupted.

Two copies of complete system and data backup are created each weekend. One of these copies is stored in a secure room at the Maple Grove location. The second copy is stored in a secure room at the Woodbury location. These backups are stored indefinitely. Incremental backups of all files on the network are made each day. The incremental backups are kept for 6 weeks.

DRC utilizes a storage area network (SAN) for maximum speed, flexibility, and redundancy in our data storage solution. Servers are connected to the SAN via redundant connections to ensure minimum interruptions due to hardware failures. The SAN allows disk space to be reallocated with ease for availability to those applications or servers as needed. The SAN currently houses 13 Terabytes of storage and is expandable to 26 Terabytes.

Chapter Nine: Summary Demographic, Program, and Accommodation Data for the 2006 Writing PSSA

ASSESSED STUDENTS

The total number of answer documents processed by grade level for the 2006 PSSA Writing assessment is presented on the first line of Table 9-1. Also shown is the number and percent of students with PSSA scores in writing, followed by those not having a score. Assessed students include those from public schools who are required to participate as well as those from a small number of non-public schools (fewer than 1,000 per grade level) that elected to participate. Also included were home-schooled students that numbered fewer than 100 per grade.

Table 9-1. Students Assessed on the 2006 PSSA

	Grade 5		Grade 8		Grade 11	
	Number	Percent	Number	Percent	Number	Percent
Number of answer documents processed	135,119		148,257		139,758	
Students with a writing score	131,463	97.3	142,960	96.4	132,210	94.6
Number processed but not assessed (without a total score)	3,656	2.7	5,297	3.6	7,548	5.4

As may be observed from Table 9-1, not all students were assessed. Although there are a variety of reasons for this, the major ones pertain to (1) excusal due to significant cognitive disability, (2) absenteeism, and (3) a situation in which there was a non-attempt on the part of the student and no exclusion code was marked by school personnel. The number of students without scores for these three reasons is presented in Table 9-2.

Students in an assessed grade who met each of the following criteria were excused from the PSSA: (1) had a significant cognitive disability, (2) required intensive instruction, (3) required adaptation and support to perform or participate meaningfully, (4) required substantial modification of the general education curriculum, (5) participation in the general education curriculum differed markedly in form and substance from that of other students (see PSSA Handbook for Assessment Coordinators and Administrators: Grades 5, 8, and 11 Writing, PDE, February, 2006, p. 7). Instead, these students participated in an alternate writing assessment. Two categories of absenteeism, (1) extended absence from school that continued beyond the assessment window and (2) being absent without makeup for at least one section of the writing assessment is combined to form a single absent category in Table 9-2. Non-attempt categorization pertains to a situation in which a student did not meet the criteria for having attempted one or more of the sections of the writing assessment and no exclusion code was marked.

Table 9-2. Counts of Students without Scores on the 2006 PSSA

Reason for Non-Assessment	Grade 5		Grade 8		Grade 11	
	N	Pct	N	Pct	N	Pct
Alternate Assessment for writing	831	22.7	963	18.2	946	12.5
Absent Writing	619	16.9	1,521	28.7	2,849	37.7
Non-Attempt Writing	1,843	50.4	2,303	43.5	2,773	36.7

COMPOSITION OF SAMPLE USED IN SUBSEQUENT TABLES

The results presented in the subsequent tables are based on those students who have a score in writing. Analyses were conducted using the individual student data file of July 26, 2006. Because some student file updates may occur subsequent to these analyses, there could be small differences in the counts although percentages would likely differ by only a fraction of a percentage point.

COLLECTION OF STUDENT DEMOGRAPHIC INFORMATION

Data for these analyses were obtained primarily from information supplied by school district personnel through the DRC Student Precode System, a multi-phase process by which student data may be imported, verified, corrected, and updated. Some data such as accommodation information is marked directly on the student answer document at the time the PSSA is administered.

DEMOGRAPHIC CHARACTERISTICS

Frequency data for each category is presented in Table 9-3. Percentages are based on all students with a score in writing as shown at the bottom of the table.

Table 9-3. Demographic Characteristics of 2006 PSSA

Demographic or Educational Characteristic	Grade 5		Grade 8		Grade 11	
	Number	Percent	Number	Percent	Number	Percent
Gender						
Female	64,278	48.9	69,512	48.6	65,412	49.5
Male	66,990	51.0	73,147	51.2	66,480	50.3
Race/Ethnicity						
American Indian or Alaskan Native	168	0.1	246	0.2	215	0.2
Asian or Pacific Islander	3,344	2.5	3,123	2.2	3,146	2.4
Black/African American non-Hispanic	20,204	15.4	21,503	15.0	15,454	11.7
Latino/Hispanic	8,341	6.3	7,999	5.6	5,203	3.9
White non-Hispanic	98,355	74.8	109,152	76.4	107,157	81.1
Multi-Racial/Ethnic	791	0.6	551	0.4	643	0.5
Educational Category and Other Demographic Groups						
IEP (not gifted)	20,293	15.4	20,809	14.6	15,961	12.1
Student exited IEP in last 2 years	892	0.7	373	0.3	239	0.2
Gifted and has an IEP	6,685	5.1	8,549	6.0	7,502	5.7
504 Plan / Chapter 15	1,030	0.8	1,040	0.7	813	0.6
Title I	38,454	29.3	26,799	18.7	14,874	11.3
Title III (3 categories below)						
Served	1,863	1.4	1,480	1.0	757	0.6
Not Served	8,294	6.3	9,043	6.3	8,901	6.7
Formerly served (2 yr monitoring)	665	0.5	486	0.3	291	0.2
Migrant Student	377	0.3	317	0.2	197	0.1
LEP (not 1st year of enrollment)	2,793	2.1	2,054	1.4	1,171	0.9
LEP in 1st yr of enrollment	227	0.2	176	0.1	203	0.2
Exited ESL/bilingual program within last 2 yrs	1,022	0.8	756	0.5	465	0.4
Foreign Exchange Student	13	0.0	5	0.0	127	0.1
Economically Disadvantaged	47,416	36.1	46,569	32.6	30,034	22.7
Hurricane Katrina displacement	9	0.0	8	0.0	10	0.0
Enrollment						
Current Enrollment in school of residence after Oct 1, 2005	4,421	3.4	4,635	3.2	3,420	2.6
Current Enrollment in district of residence after Oct 1, 2005	2,954	2.2	3,286	2.3	2,716	2.1
Current Enrollment as PA resident after Oct 1, 2005	1,108	0.8	1,182	0.8	960	0.7
Current Enrollment in district of residence after Oct 1, 2003	18,191	13.8	15,849	11.1	11,103	8.4
Enrolled in district of residence after Oct 1, 2002 but before Oct 1, 2003	5,880	4.5	6,164	4.3	6,816	5.2
Homeless as defined by McKinney-Vento Act	176	0.1	175	0.1	66	0.0
Number Scored	131,463		142,960		132,210	

EDUCATION IN NON-TRADITIONAL SETTINGS

For each category the number and percent are presented for all students with a score in writing. Table 9-4 reveals an incidence of less than one percent for the majority of these settings. Also shown are home schooled students assessed by parental request.

Table 9-4. Participation in 2006 PSSA by Students in Non-Traditional Settings

Non-Traditional Educational Settings	Grade 5		Grade 8		Grade 11	
	Number	Percent	Number	Percent	Number	Percent
Court/agency placed	146	0.1	538	0.4	669	0.5
Homebound instruction	6	0.0	10	0.0	14	0.0
Special education student placed in program outside the district of residence	46	0.0	127	0.1	138	0.1
Special education student placed in program located in one building/site within the district of residence	390	0.3	90	0.1	52	0.0
Student placed in Approved Public Alternative Education Program	188	0.1	329	0.2	398	0.3
Special education student placed in Approved Public Alternative Education Program	13	0.0	55	0.0	48	0.0
Student placed in Approved Private School (APS)	108	0.1	217	0.2	219	0.2
Student attends an intermediate unit (IU) program/classroom	246	0.2	468	0.3	411	0.3
Home schooled student assessed by parental request	52	0.0	53	0.0	30	0.0

PARTICIPATION IN PSSA BY CAREER AND VOCATIONAL EDUCATION STUDENTS

Table 9-5 summarizes the total number of students receiving a score on the PSSA who were enrolled in an approved CTE program. Some of these students are dually coded as enrolled in a tech prep program.

Table 9-5. Participation by Grade 11 Career and Vocational Education Students

Career and Vocational Education Categories	Number	Percent of CTE Students	Percent of all Assessed Students
Students enrolled in a CTE program approved by Career & Technical Education System	13,396	80.3	10.1
Students enrolled in a tech prep program who are dually coded as CTE	3,280	19.7	2.5
Number scored classified as CTE	16,676	100	12.6
Students enrolled in a tech prep program but NOT dually coded as CTE	764		0.6
Number Scored classified as CTE or as tech prep only	17,440		13.2

Table 9-6 provides data regarding the type of school setting in which the grade 11 students

receive their career and vocational education. Table 9-6 also presents information regarding the student's career cluster. In this table the totals are based on the summation of assessed students across type of school settings and across program areas. The associated percents relate to the total numbers of career and vocational education (CTE) students with a score in writing.

Table 9-6. School Setting and Career Cluster Categories of Grade 11 CTE Students

Student attends:	**Number**	**Percent**
Comprehensive CTE for full day	2,228	13.4
District High School with CTE classes	3,499	21.0
Charter School with an Approved CTE	104	0.6
Career and Technical Center part time	9,187	55.1
Not coded	1,658	9.9
Totals	16,676	100 %
Career Cluster in which the student is classified:		
Agriculture	935	5.6
Architecture and Construction	2,628	15.8
Arts and Communication	803	4.8
Business Management	998	6.0
Education and Training	399	2.4
Finance	40	0.2
Government and Public Administration	11	0.1
Health Science	1,292	7.7
Hospitality and Tourism	937	5.6
Human Services	1,455	8.7
Information Technology	748	4.5
Law and Public Safety	394	2.4
Manufacturing	1,052	6.3
Marketing, Sales and Service	723	4.3
Science and Technology	696	4.2
Transportation and Logistics	1,903	11.4
Not coded	1,662	10.0
Totals	16,676	100 %

PRIMARY DISABILITY OF IEP STUDENTS ASSESSED ON THE PSSA

School personnel supplied the primary disability information for those students who had an IEP (not gifted) through the DRC Student Precode System. Beginning with 2006, the disability categories are presented in a sequence that matches a Department of Education numbering system and two previously separate categories were combined. In Table 9-7, for each disability category, the number and percent are presented for all students with a score in writing coded with a disability. For example, if 20,000 students statewide had a coded disability and 10,000 students were classified as having a specific learning disability, the table entries will show 10,000 followed by 50%. Uniformly, specific learning disability is the category with the highest incidence of occurrence. The last row of Table 9-7 presents the percent of all assessed students who have a coded primary disability.

Table 9-7. Incidence of Primary Disabilities Among IEP Students Assessed on the 2006 PSSA

Primary Disability of Students Having an IEP	Grade 5		Grade 8		Grade 11	
	Number	Percent	Number	Percent	Number	Percent
Traumatic Brain Injury	29	0.2	43	0.2	33	0.2
Hearing Impairment incl. Deafness	181	0.9	187	1.0	156	1.1
Specific Learning Disability	11,625	60.9	13,695	70.9	10,718	73.8
Mental Retardation	700	3.7	985	5.1	802	5.5
Orthopedic Impairment	40	0.2	38	0.2	33	0.2
Emotional Disturbance	1,400	7.3	2,137	11.1	1,813	12.5
Speech or Language Impairment	3,629	19.0	862	4.5	202	1.4
Visual Impairment incl. Blindness	48	0.2	52	0.3	46	0.3
Deaf/Blind	8	0.0	11	0.0	7	0.0
Multiple Disabilities	109	0.6	64	0.3	45	0.3
Autism	348	1.8	283	1.5	108	0.7
Other Health Impairment	977	51	957	5.0	553	3.8
Number Scored	19,094	100	19,314	100	14,516	100
Percent of Total Assessed Students with a Coded Disability		14.5		13.5		11.0

TEST ACCOMMODATIONS PROVIDED

School personnel supplied information regarding accommodations of various types that a student may have received while taking the PSSA. These included changes in test environment, modified test formats, and special arrangements and assistive devices. The frequency with which these accommodations were utilized is summarized in Tables 9-8, 9-9, and 9-10. The values in the table are based on all students with a score in writing. Please note that a glossary of accommodation terms as applied to the PSSA is provided in Table 9-13 at the end of this chapter.

CHANGES IN TEST ENVIRONMENT

There were seven categories of test environment changes on the 2006 PSSA writing assessment. As depicted in Table 9-8, the most common accommodations were small group testing, testing in a separate room, scheduled extended time and requested extended time.

Table 9-8. Incidence of Changes in Test Environment on the 2006 PSSA

Type of Change in Test Environment	Grade 5		Grade 8		Grade 11	
	Number	Percent	Number	Percent	Number	Percent
Scheduled Extended Time	6,835	5.2	4,831	3.4	4,230	3.2
Requested Extended Time	2,697	2.1	3,441	2.4	3,045	2.3
Separate Room	8,763	6.7	6,298	4.4	4,250	3.2
Hospital/Home Testing	24	0.0	74	0.1	83	0.1
Multiple Test Sessions	1,688	1.3	1,515	1.1	859	0.6
Small Group Testing	11,116	8.5	9,966	7.0	6,740	5.1
Other	426	0.3	197	0.1	156	0.1

MODIFIED TEST FORMATS

There were five categories of test format modifications in the 2006 PSSA writing assessment. As depicted in Table 9-9, the actual frequencies are quite low, generally representing less than a tenth of one percent of assessed students statewide.

Table 9-9. Incidence of Test Format Modifications on the 2006 PSSA

Type of Test Format Modification	Grade 5		Grade 8		Grade 11	
	Number	Percent	Number	Percent	Number	Percent
Braille Edition	7	0.0	8	0.0	10	0.0
Large Print Edition	77	0.1	66	0.0	44	0.0
Word Processor	83	0.1	64	0.0	130	0.1
Signed Version	20	0.0	14	0.0	20	0.0
Other	19	0.0	31	0.0	45	0.0

SPECIAL ARRANGEMENTS/ASSISTIVE DEVICES

On the 2006 PSSA writing assessment, there were eight categories of accommodations in the form of special arrangements or assistive devices. The frequency with which these accommodations were utilized is summarized in Table 9-10. The actual frequencies are quite low, generally representing less than four-tenths of one percent of assessed students statewide.

Table 9-10. Incidence of Special Arrangements/Assistive Devices on the 2006 PSSA

Type of Arrangement or Assistive Device	Grade 5		Grade 8		Grade 11	
	Number	Percent	Number	Percent	Number	Percent
Braille Writer	4	0.0	3	0.0	11	0.0
Test Administrator transcribed illegible writing	478	0.4	193	0.1	127	0.1
Interpreter signed directions	39	0.0	45	0.0	55	0.0
Magnification device	6	0.0	15	0.0	4	0.0
Test administrator marked test at student direction (MC only)	76	0.1	47	0.0	28	0.0
Typewriter, word processor or computer	256	0.2	321	0.2	303	0.2
Qualified interpreter for LEP student	106	0.1	85	0.1	62	0.0
Other	179	0.1	245	0.2	83	0.1

THE INCIDENCE OF ACCOMMODATIONS AND IEP AND LEP STATUS

It is reasonable to expect that students with an IEP would receive the majority of accommodations; however, certain accommodations are specific to particular disabilities or to students classified as limited English proficient (LEP). A cross-tabulation between each of the accommodations and IEP and LEP status revealed a much greater incidence for the categorical students. This is most clearly depicted in the frequently occurring accommodations. To illustrate, several of these results were selected for display in Table 9-11.

For the IEP analysis, the column headings refer to students classified as IEP (IEP) and non-IEP (NIEP). In each instance there is a considerably larger percent of IEP students receiving the accommodation than NIEP students. There is a general tendency to observe a decrease in the percentage of IEP students receiving these accommodations in the progression from lower to higher grade levels.

The analysis for students with limited English proficiency was based on the formation of a new variable by combining two separate items dealing with a student's LEP status. The two items differentiated between those LEP students who were in their first year of enrollment in U.S. schools and those who were not. The constructed variable, labeled LEPC in Table 9-11, was assigned a value of one if either of the two items was marked and was zero otherwise. Non-LEP is labeled as NLEPC. The accommodations most frequently received by LEPC students are presented. In each instance there is a considerably larger percent of LEPC students receiving the accommodation than NLEPC students. As is typical there is a decrease in the percentage of LEPC students receiving these accommodations in the progression from lower to higher grade levels except the use of a qualified interpreter, which remained at essentially the same rate.

Table 9-11. Percent of IEP and LEP Students Receiving Selected Accommodations

	Grade 5		Grade 8		Grade 11	
Accommodation Received	**IEP**	**NIEP**	**IEP**	**NIEP**	**IEP**	**NIEP**
Scheduled extended time	25.5	1.5	18.8	0.7	21.5	0.7
Test in separate room	34.7	1.6	26.0	0.7	23.0	0.5
Test in small group setting	44.7	1.8	41.3	1.1	36.9	0.7
Accommodation Received	**LEPC**	**NLEPC**	**LEPC**	**NLEPC**	**LEPC**	**NLEPC**
Scheduled extended time	13.3	5.0	9.6	3.3	9.4	3.1
Test in separate room	16.3	6.4	12.0	4.3	10.7	3.1
Test in small group setting	19.6	8.2	16.1	6.8	10.8	5.0
Qualified interpreter for LEP student	3.1	0.0	3.2	0.0	3.2	0.0

THE INCIDENCE OF ACCOMMODATIONS AND PRIMARY DISABILITY CLASSIFICATION

To further delineate the use of commonly employed accommodations, a grade level breakdown by major primary disability is presented in Table 9-12. A selection was made based on the more frequently occurring categories of disability and accommodations rather than displaying data for all of them. As may be seen from a perusal of Tables 9-8, 9-9, and 9-10, the accommodations with the larger frequencies are those that involve a change in test environment or that necessitate special arrangements. Selected for incorporation in Table 9-12 are the five test environment accommodations with frequencies in excess of 1,000 at all grade levels and the four special arrangement accommodations with the largest frequencies at grade 5. Accommodations concerned with test format modifications tended to be highly specific to particular and infrequent

disability categories or to students classified as limited English proficient (LEP) and were not included in Table 9-12. Seven Primary Disability categories were selected that had a minimum of 100 students so classified at each grade level.

The entries for Table 9-12 represent the number and percent of students with a particular disability (columns) who received the listed accommodation (rows). For example, if 200 students out of 500 classified with a particular disability received scheduled extended time, the table entries will show 200 followed by 40%.

The most prominent and consistent findings from Table 9-12 are (1) the heavy use of scheduled extended time, a separate room, and small group settings for all disability categories except speech and language impairment and that (2) in each instance the percent of 5th grade students receiving these three accommodations exceeded that of 8th and 11th grade students by about 10 to 20 percent.

Table 9-12. Incidence of Test Accommodations Received for Selected Primary Disability Classifications on the 2006 PSSA

		Primary Disability of Assessed Student with an IEP: Number and Percent													
Type of Accommodation Received	Grade Level	Autism		Emotional Disturbance		Deafness / Hearing Impairment		Mental Retardation		Other Health Impairment		Specific Learning Disability		Speech or Language Impairment	
Scheduled extended time	5	95	27%	383	27%	19	10%	248	35%	285	29%	3405	29%	145	4%
	8	62	22%	509	24%	35	19%	270	27%	162	17%	2324	17%	18	2%
	11	30	28%	475	26%	44	28%	211	26%	112	20%	2057	19%	7	4%
Student-requested extended time	5	12	3%	52	4%	4	2%	26	4%	31	3%	445	4%	78	2%
	8	21	7%	89	4%	6	3%	270	27%	39	4%	735	5%	16	2%
	11	3	3%	64	4%	5	3%	28	4%	22	4%	430	4%	8	4%
Separate room	5	154	44%	442	32%	35	19%	287	41%	424	43%	4812	41%	221	6%
	8	70	25%	530	25%	30	16%	320	32%	299	31%	3710	27%	35	4%
	11	27	25%	433	24%	26	17%	242	30%	140	25%	2459	23%	12	6%
Multiple test sessions	5	33	10%	154	11%	1	1%	78	11%	58	6%	760	6%	33	1%
	8	28	10%	230	11%	10	5%	64	6%	32	3%	513	4%	3	0%
	11	1	1%	183	10%	11	7%	38	5%	13	2%	249	2%	1	1%
Small group testing	5	184	53%	593	42%	38	21%	396	57%	517	53%	6257	54%	252	7%
	8	126	44%	959	45%	54	29%	460	47%	433	45%	5602	41%	49	6%
	11	48	44%	771	42%	46	30%	354	44%	206	37%	3832	36%	19	9%
Typewriter, word processor or computer	5	27	8%	11	1%	2	1%	2	0%	27	3%	98	1%	13	0%
	8	16	6%	21	1%	1	0%	8	1%	34	4%	150	1%	3	0%
	11	12	11%	19	1%	1	1%	12	2%	15	3%	166	2%	1	0%
Test admin. transcribed illegible writing	5	19	6%	16	1%	2	1%	12	2%	26	3%	218	2%	22	1%
	8	9	3%	14	1%	0	0%	8	1%	12	1%	84	1%	1	0%
	11	4	4%	2	0%	1	1%	2	0%	4	1%	62	1%	1	0%

Note: Results displayed are for most frequently occurring accommodations and disability classifications

GLOSSARY OF ACCOMODATIONS TERMS

Table 9-13 provides a brief description of accommodations terms as utilized in the PSSA. School personnel identified the accommodations that a student received by marking a bubble in the student answer document as seen in the left column of Table 9-13. The right column contains an explanation abstracted from the *2006 Accommodations Guidelines* (PDE, December, 2005, see especially pages 4-14).

Table 9-13. Glossary of Accommodations Terms as Applied in the PSSA

Type of Testing Accommodation	**Explanation**
Student was given the following changes in test environment *(mark all that apply)*	
Scheduled extended time	Extended time may be allotted for each section of the test to enable students to finish.
Student-requested extended time	A student may request extended time if working productively.
Tested in a separate room	A separate room may be used to reduce distraction.
Hospital/home testing	A student who is confined to a hospital or to home during the testing window may be tested in that environment.
Multiple test sessions	Multiple test sessions may be scheduled for the completion of each test section; however, a test section must be completed within one school day.
Small group testing	Some students may require a test setting with fewer students or in a setting apart from all other students.
Other	Other accommodations may be appropriate and available if they do not compromise the integrity of the assessment. Questions may be directed to PDE.
Student used the following modified test format(s) *(mark all that apply)*	
Braille edition	Students may use a Braille edition of the test. Answers must then be transcribed into the answer booklet without alteration.
Large print edition	Students with visual impairments may use a large print edition. Answers must then be transcribed into the answer booklet without alteration.
Word processor	Students with an identified need may use a word processor or a typewriter. Answers must then be transcribed into the answer booklet without alteration.
Signed version	Qualified interpreters may sign directions for all assessments. Mathematics questions may be signed. On the writing assessment only the writing prompt may be signed. Signing the passage and/or questions on the reading test is not permitted.

Other	Other accommodations may be appropriate and available if they do not compromise the integrity of the assessment. Questions may be directed to PDE.
Student used the following special arrangements / assistive devices *(mark all that apply)*	
Braille writer *(with no thesaurus, spell- or grammar checker)*	Students using this device as part of their regular program may use it on the PSSA.
Test administrator transcribed illegible writing	Dictation of student's responses is not allowed, but administrator may transcribe handwriting that is extremely difficult to read.
Interpreter signed directions	Deaf/hearing impaired students may receive test directions from a qualified interpreter.
Magnification devices	Devices to magnify print may be used for students with visual impairments.
Test administrator marked test at student's direction *(multiple-choice answers only)*	A test administrator may mark an answer booklet at the direction of a student. (e.g., a student may point to a multiple-choice answer with the test administrator marking the response in the answer booklet).
Typewriter, word processor or computer *(with thesaurus, spell- or grammar-checker turned off)*	An allowable accommodation as a typing function only for students with identified need. Supports such as dictionaries, thesauri, spell checkers and grammar checkers must be turned off.
Qualified interpreter for LEP student	An interpreter may translate directions or clarify instructions for the assessments. They may translate, but not define specific words or test questions on the mathematics test. On the reading test interpreters may only translate directions and may not translate or define words in the passage or test questions.
Other	Other accommodations may be appropriate and available if they do not compromise the integrity of the assessment. Questions may be directed to PDE.

Chapter Ten: Form Analysis and Item Calibration

TEST FORM STATISTICS

Table 10-1 contains an overview of the form-level data. Test length in total number of points (L), mean number of points received (P), standard deviations (SD), test reliability (R), and traditional standard error of measurement (SEM) are shown by grade and content area. These statistics are based on the total test using both multiple-choice and open-ended tasks for the common sections of each form. Detailed item-level statistics for the common items can be found in Appendices H through M.

Test reliability can be thought of as the correlation between scores if the students could be tested twice with the same instrument without the second testing being affected by the first. It can also be conceptualized as the extent to which different items from the same domain would result in similar ordering of students. It is computed as the ratio of the variance associated with *true scores* divided by the total variance. The total variance contains a component due to the variance in true scores and a component due to the imperfections in the measurement process.

$$R = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2} = \frac{\sigma_O^2 - \sigma_e^2}{\sigma_O^2}$$

When a test is composed of subtests that require different cognitive tasks, however – operant versus respondent items, for instance – the traditional formula for Cronbach's alpha reliability may underestimate the true total test reliability (Qualls, 1995). In such cases, it can be more informative to compute the stratified alpha coefficient (Cronbach, Schönemann, & McKie, 1965), a weighted reliability coefficient that accounts for the contribution of each subtest to the overall test variance:

$$R = 1 - \frac{\sum \sigma_j^2 (1 - \alpha_j)}{\sigma_X^2}$$

where σ_j^2 is the variance of stratum (or subtest) j, α_j is the reliability of stratum j, and σ_X^2 is the variance of the test as a whole. Reliabilities presented in Chapters 10 and 13 of this technical report are stratified reliability coefficients, computed by dividing the test into three strata: multiple-choice items, Composition scores, and Revising and Editing scores. This division was chosen based on both conceptual grounds and test-component variance-covariance matrices. It should be noted that like other internal-consistency reliability coefficients, this index is not sensitive to other types of errors that can affect test scores, such as temporal stability or variability in performance that might occur across testing occasions. It is also not sensitive to rater error. Consequently, this index might be positively biased by these factors.

The reliability coefficient is a "unitless" index, which can be compared from test to test. The *standard error of measurement* is another indicator of precision. If everyone being tested had the same *true score*[2], there would still be some variation in observed scores due to imperfections

[2] True score is the score the person would receive if the measurement process were perfect.

in the measurement process, such as random differences in attention during instruction or concentration during testing. The standard error is defined as the *standard deviation*[3] of the distribution of observed scores for students with identical true scores. Because the SEM is an index of the random variability in test scores in actual score units, it represents important information for test score users.

Generally speaking, reliabilities go up with an increase in test length and population heterogeneity and go down with shorter tests and more homogeneous populations.

Table 10-1. 2006 Summary of Common Item Performances

	Writing				
Grade	**L**	**P**	**SD**	**R**	**SEM**
5	100	68.6	13.0	0.75	6.4
8	100	71.6	14.1	0.81	6.2
11	100	68.0	13.2	0.79	6.1

The standard deviation shown in the table is the standard deviation of observed scores. Assuming normally distributed scores, one would expect about two-thirds of the observations to be within one standard deviation of the mean. An estimate of the standard deviation of the true scores can be computed as $\hat{\sigma}_T = \sqrt{\sigma_x^2 - \sigma_x^2(1-\rho_{xx})}$. As an example, for grade 5, this would be $\sqrt{13.0^2 - 8.6^2} = 9.75$.

The conditional standard error of measurement (CSEM) also indicates the degree of measurement error in score units, however, does so as a function of one's actual test score. Therefore, the CSEM may be especially useful in characterizing measurement precision in the neighborhood of a score level used for decision-making—such as cut scores for identifying students who meet a performance standard. The CSEMs for Reading and Mathematics are documented in Appendix N in the column labeled "Scale Score SE."

TRADITIONAL ITEM STATISTICS

Although all items were previously reviewed for both content and statistical quality, a thorough item analysis was conducted to ensure that the items and forms performed as expected. With any psychometric model, an item analysis is a search for unexpected results. For example, *more able*[4] students are expected to pass easy items and *less able* students are expected to fail difficult items. If either of these situations does not occur, the item should be reviewed to determine the nature of the problem and the characteristics of the students affected.

The most familiar indices of item performance are *proportion correct* (P-Value) and item reliability. Reliability for dichotomous items is typically represented by the *point biserial correlation* coefficient. The correlation will have a positive value when the mean score of the

[3] The standard deviation of a distribution is a measure of the dispersion of the observations. For the normal distribution about 16% of the observations are more than one standard deviation above the mean and the same percentage are more than one standard deviation below the mean. Using the data in table 10.1, about 68% of students with true scores of 70 for grade 5 writing will have observed scores between 61.4 and 78.6.

[4] Following the Rasch literature, *ability* is used in this discussion as a generic term for the construct that is being measured by the exam. *Competence, achievement, learning* and *status* are among the alternatives that are sometimes used, but are all subject to some degree of misinterpretation.

students answering correctly is higher than the mean score of the students answering incorrectly. This indicates that students who did well on the total test tended to do well on this item. The index will take its maximum theoretical value of 1.0 if *every* student who answered the item correctly scored better on the test than *any* student who answered incorrectly[5].

The P-Value is a subtler indicator of item quality. If there is a *more able* way to miss an item, the item will appear more difficult than expected. Conversely, if there is a *less able* way to pass the item, it may appear surprisingly easy.

P-values for constructed response items were obtained by dividing the mean points by the total number of possible points. While the CR P-values are on the same scale as the MC P-values, unlike the MC P-values they cannot be interpreted as the proportion of students answering the item correctly. Otherwise, the interpretation of CR P-values is consistent with the interpretation of MC P-values, with higher values indicating easier items.

Table 10-2 provides some distributional indices for the P-Value and point biserial correlation (PtBis) for the multiple-choice items on the common form in each grade and content area.

In general, with the mean P-Values in the range of 0.75 – 0.79, the PSSA was reasonably challenging to most students. With the average point biserial correlations ranging from .34 to .37, the overall item quality was good. It should be noted that rules of thumb for interpreting these statistics should be flexible relative to the purposes and uses of test scores. An average P-Value around 0.65 (or slightly higher) is considered advantageous for spreading out students. Similarly, point-biserial correlations are often grouped (e.g., above 0.20 being "adequate," above 0.30 being "good," and above 0.40 being "excellent"). However, in the context of a criterion-referenced testing program, the best items for covering content domains and depth-of-knowledge levels do not always fall within these guidelines.

[5] It is legitimate to view the point biserial correlations as standardized means. A positive value means students who chose that response had a higher mean score than the average student; a negative value means students who chose that response had a lower than average mean score.

Table 10-2. Common Form Statistics by Grade and Content for Multiple-Choice Items

	Writing	
Grade 5	**P-Value**	**PtBis**
Average	0.77	0.34
Minimum	0.51	0.21
Maximum	0.92	0.43
Median	0.79	0.35
	Writing	
Grade 8	**P-Value**	**PtBis**
Average	0.75	0.34
Minimum	0.42	0.26
Maximum	0.88	0.41
Median	0.80	0.33
	Writing	
Grade 11	**P-Value**	**PtBis**
Average	0.79	0.37
Minimum	0.49	0.28
Maximum	0.93	0.45
Median	0.84	0.38

RASCH ITEM STATISTICS AND EQUATING

WINSTEPS© software implementing the Rasch model was used to obtain estimates of logit difficulties for both dichotomously- and polytomously-scored items. The parameters estimated for polytomous items are the *step difficulties* associated with the Masters Partial Credit model. This software is capable of handling all the item types currently in use with the PSSA. WINSTEPS© version 3.54 was used for all calibrations. See Wright and Masters (1982) and Rasch (1960) for further information about the models used for these analyses.

The Rasch model expresses item difficulty (and student ability) in units referred to as *logits*, rather than in percent correct. In the simplest case, a logit is a transformed P-Value with the average P-Value becoming a logit of zero. In this form, logits resemble z-scores or standard normal deviates; a very difficult item might have a logit of +4 and a very easy item might have a logit of –4. However, they have no formal relationship to the normal distribution.

The logit metric has several mathematical advantages over P-Values. It is an interval scale, meaning that two items with logits of zero and one (respectively) are the *same distance* apart as items with logits of +3 and +4. Logits are not dependent on the ability level of the students. For example, a form can have a mean logit of zero, whether the average P-Value of the sample is 0.8 or 0.3.

The standard Rasch calibration procedure arbitrarily sets the mean difficulty of the items on any form at zero. Under normal circumstances where all students are administered a common set of

items, any item with a P-Value lower than the average item on the form receives a positive logit difficulty and any item with a P-Value higher than the average receives a negative logit. Consequently, the logits for any calibration, whether it is a third grade reading test or a high school science test, relate to an arbitrary origin defined by the center of items on that form. The average third grade reading item will have a logit of zero; the average high school science item will have a logit of zero. Logits for both item difficulties and student abilities are placed on the same scale and relate to the same mean item difficulty.

There are any number of other arbitrary choices that could be made for centering the item difficulties. Rather than using all the items, the origin could be defined by a subset. For the PSSA, all test forms in a particular grade and content area share a common block of items. The items on all forms can then be easily adjusted to a single (but still arbitrary) origin by defining the origin as the mean of the **common** items. With this done, the origins for all the forms will be statistically equal. Items on forms A and F that are equally difficult will now have *statistically* equal logit difficulties.

Note that test forms were spiraled within classrooms. In effect, students are administered the exact same set of common items but different field test or matrix sets. As a result, there are cross checks that are made to ensure the calibrations and links are reasonable across forms. The goal of spiraling is to achieve a *statistically equivalent* sample of students across forms with equal standard deviations and arbitrary means. Any differences in performance observed among the groups should be due only to differences in form difficulty. After linking, the mean of the logit abilities should be statistically equal for each sample of students.

Winsteps' Outfit (outlier-sensitive fit) index is sensitive to outliers—e.g., aberrant responses to items with difficulty far from a person's ability—and indicates overfit for imputed responses and underfit for lucky guesses and careless mistakes. Outfit values for items are reported beginning in Appendix E. Here, Outfit is expressed on a standardized metric (*t*), which is more oriented toward statistical significance. Specifically, *t* shows the degree of improbability in the data (i.e., its statistical significance) if the data actually did fit the model. The expected value is 0.0 with values significantly less than 0.0 indicating too much predictability and values significantly greater than 0.0 indicating lack of predictability.

Because of the equivalent samples, the common items should have the same P-Values regardless of which form and sample is being considered. Finally, for all items, a plot of the relationship between the P-Value and the logit should fall along a single, curved line. Figure 10.1 through 10.3 plot this relationship for common multiple-choice items. The curves are nearly linear in the center, but curve towards asymptotes of one and zero, respectively, on the left and on the right. The graphs show that items with low P-values (indicating a more difficult item that fewer students answered correctly) also showed higher logit difficulty, and items with high P-values had lower logit difficulties. The spread of the graph points is indicative of the dispersion of item difficulties in the common items.

Figure 10.1. 2006 Grade 5 Writing Logit Difficulties versus P-Values

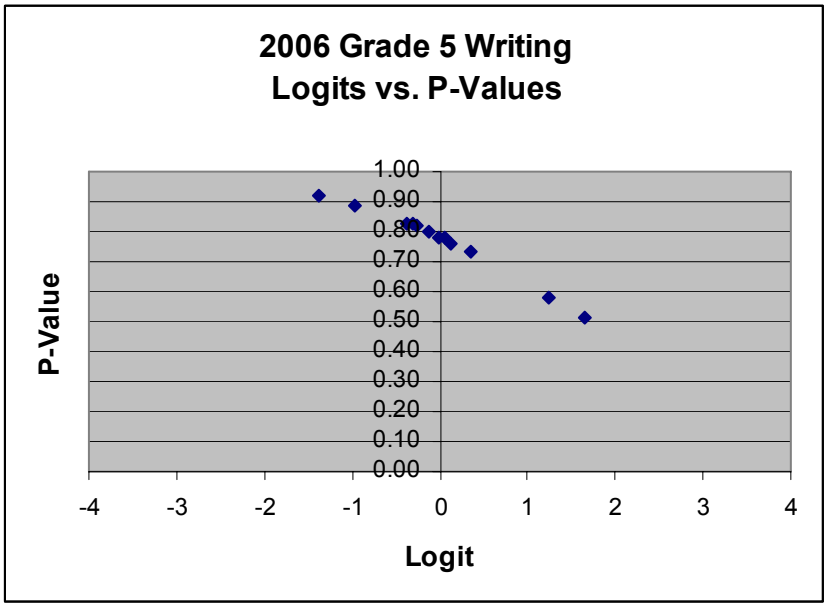


Figure 10.2. 2006 Grade 8 Writing Logit Difficulties versus P-Values

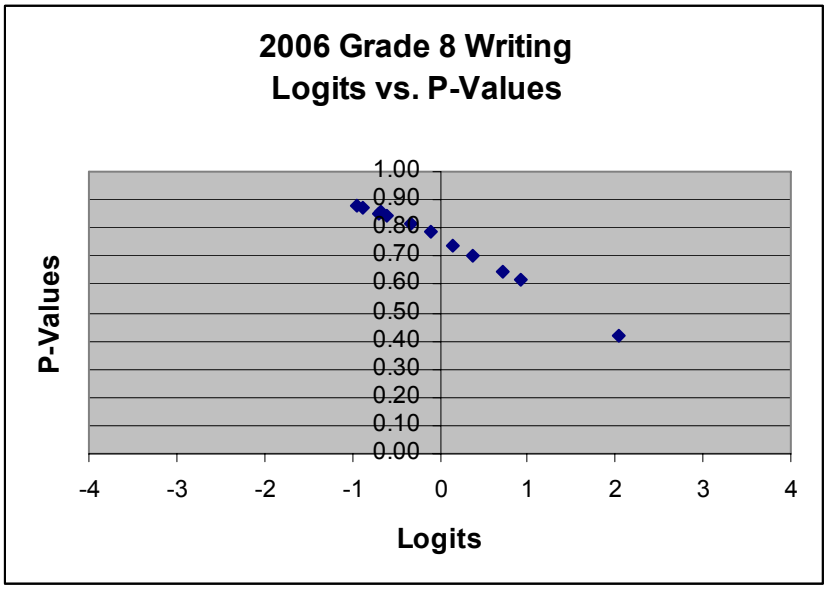
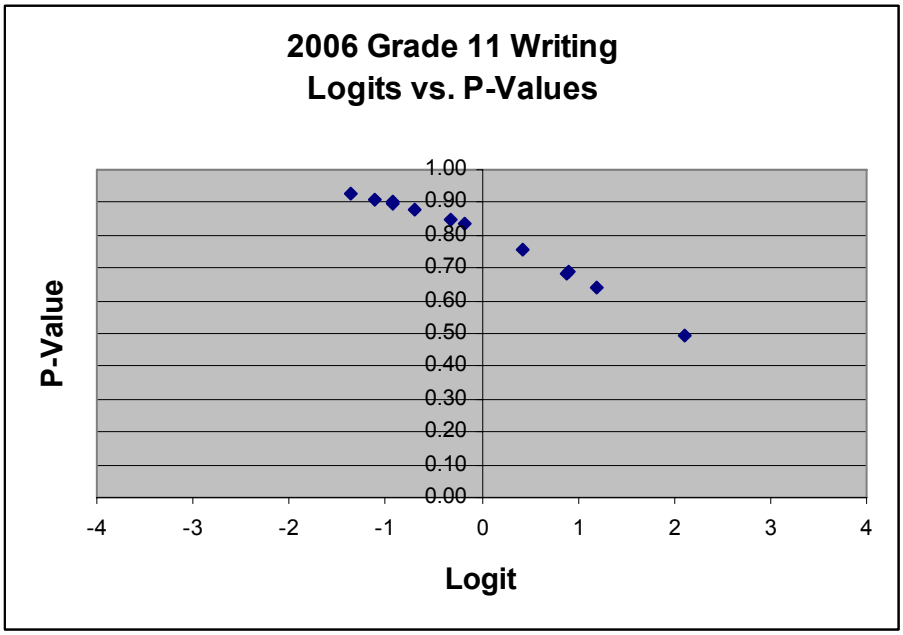


Figure 10.3. 2006 Grade 11 Writing Logit Difficulties versus P-Values



Below are the mean P-Values by form for the common multiple-choice items. The extent to which the mean P-values across forms are similar indicates the extent to which the student populations taking each form are of approximately of equal ability. This equivalence of ability distributions across forms is the desired outcome of spiraling and allows for optimum analysis of the embedded field test items.

Grade 5 Writing		
Form	**Mean P-Value**	**Std. Dev.**
1[6]	0.77	0.12
2	0.77	0.12
3	0.78	0.11
4	0.77	0.12
5	0.77	0.12
6	0.77	0.12
7	0.78	0.11
8	0.77	0.11
9	0.77	0.12
10	0.77	0.13
Avg	**0.77**	**0.12**

Grade 8 Writing		
Form	**Mean P-Value**	**Std. Dev.**
1	0.75	0.14
2	0.75	0.14
3	0.75	0.14
4	0.75	0.14
5	0.75	0.14
6	0.75	0.14
7	0.75	0.14
8	0.75	0.14
9	0.75	0.14
10	0.76	0.14
Avg	**0.75**	**0.14**

Grade 11 Writing		
Form	**Mean P-Value**	**Std. Dev.**
1[7]	0.79	0.14
2	0.79	0.14
3	0.79	0.14
4	0.78	0.14
5	0.79	0.14
6	0.79	0.13
7	0.79	0.12
8	0.79	0.14
9	0.79	0.14
10	0.79	0.13
Avg	**0.79**	**0.14**

[6] For both reading and mathematics in all grades, form 1 was used to generate modified versions (e.g., Large Print and Braille) of the common form; thus, the mean P-Values for these forms are somewhat lower.

[7] In all grades, form 1 was used to generate modified versions (e.g., Large Print and Braille) of the common form; thus, the mean P-Values for these forms are somewhat lower.

Chapter Eleven: Linking

Because the Writing assessment was new in 2006 for grades 5 and 8 and the Grade 11 test has changed substantially since the last administration—including the setting of new cut scores—no linking analyses were performed.

For future linking designs, core and linking item sets will be drawn from the previous year's matrix field test. Within-year linking will be done by common calibration of the forms; current-to-bank linking will be done using prompts from the bank; and current-to-previous-year linking will be done using matrix-to-matrix and matrix-to-core item sets.

Chapter Twelve: Scaled Scores & Performance Levels

COMMON ITEMS AND MATRIX SAMPLED ITEMS

Beginning with the design changes implemented for the 2000 PSSA, student-level scores were based on the common items only. This ensures that any decision made about students will be done in the most equitable manner. School-level scaled scores for the content areas are based on the mean of the student-level scaled scores. This ensures that the scaled scores used for school accountability directly reflect the student-level results. It is a simple matter to aggregate up to the school, district, and state levels.

For the purpose of providing school-level results at the content standard (academic standards category) level, all items on all matrix forms plus the common items are utilized. This ensures that decisions about potential school-level strengths and weaknesses are based on broad sampling of the curriculum.

SCALED SCORES FOR CONTENT STANDARDS

As of 2003, school-level scaled scores are no longer reported for the academic content standards (academic standards categories). Instead, school results are presented as the percent of total points achieved as compared to district and state level results.

INTERPRETING SCALED SCORES AND PERFORMANCE LEVELS

A *Scaled Score*, in the simplest sense, is a transformed number correct score[8]. When all students take the same items, as in the common sections of the PSSA, the more points the student earns, the higher the associated scaled score. The value of switching to the more abstract scaled score metric lies in the achievement of a more general and equitable result.

To illustrate, a raw score of 30 is meaningless unless the reader is also told how many points were possible. The same score has quite different meanings if it is based on a thirty-item test as opposed to a sixty-item test. *Number correct scores are transformed to percent correct scores to remove the effect of test length.* In the same way, a score based on sixty *difficult* items is quite different from the same score based on sixty *easy* items. *Number correct scores are transformed to scaled scores to remove the effects of test length and item difficulty.* As a result, scale scores lend themselves to interpretations at what is referred to as an interval level, while raw scores do not. Interval-level scales allow one to interpret a scale score difference of 5 points the same whether the scores are 1295 vs. 1300 or 1445 vs. 1450. Raw score differences, in this context, cannot be interpreted in this manner and are thus neither generalizable nor equatable.

The scale for the new PSSA writing assessment was established by setting the mean at all three grade levels to the 2005 proficient score cutpoint of 1236. The standard deviation was set to 100 and the scale minimum to 700 in order to create a scale similar to the Grade 11 scale.

These values are arbitrary; they could have been zero and one, or 100 and 110, or any other ordered pair without affecting any of the relationships among schools, years, students, or items. Changing the scale would simply be changing the labels on the axis of a graph without moving

[8] This is done in two steps. First, a nonlinear transformation is used to convert number correct scores to logits, and then a linear transformation is used to convert logits to scaled scores.

any of the points. Like the temperature scales of Fahrenheit and Celsius, the new scale will acquire meaning to users only with experience.

Raw to scale score tables for the Spring 2006 assessment can be found in Appendix N.

PSSA PERFORMANCE LEVELS FOR WRITING

Performance levels are another way to attach meaning to the scaled score metric. They associate precise quantitative ranges of scaled scores with verbal, qualitative descriptions of student status. While much less precise, the qualitative description of the levels is one way for parents and teachers to interpret the student scores. They are also useful in assessing the status of the school.

The current performance level descriptors, as developed by PDE and teacher panels, are given in Appendix O.

The quantitative definition of the performance levels, established through the Performance Levels Validation process, is shown in Chapter 14.

Chapter Thirteen: Test Validity and Reliability

CALIBRATION

In order to expedite the analysis process, a sample of students was selected for use in calibrating items. The sample was aimed to cover roughly 50% of the student population while preserving ethnic representation. This was done using random sampling without replacement at the district level for approximately 85% of the sample and at the school level for Pittsburgh and Philadelphia districts for approximately 15% of the sample based on 2005-2006 enrollment counts.

VALIDITY

As noted 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 tests" (AERA, APA, & NCME, 1999, p. 9). Thus, the validity of the PSSA must be judged in relation to its primary purposes as delineated in Chapter 1. Validity evidence related to test content is presented in terms of how the 2005 PSSA assessments were assembled to reflect the state content standards (more information on this is presented in Chapter 3).

The PDE commitment to validity is also evidenced by the fact that the Pennsylvania State Board of Education commissioned an independent study of an earlier version of the PSSA. That study, conducted by HumRRO, included an extensive evaluation of the items (Thacker and Dickinson, 2004) and of statistical relationships of the PSSA, including convergent and discriminant validity (Thacker, Dickinson and Koger, 2004).

RELIABILITY

This chapter provides reliability indices and standard error of measurement (SEM) for the 2006 PSSA assessments. For the Rasch model, raw scores are sufficient statistics for abilities and scale scores; performance levels set on scale scores are identical to those based on raw scores.

RELIABILITY INDICES

Stratified alpha coefficient (Cronbach, Schönemann, & McKie, 1965), a weighted reliability coefficient that accounts for the contribution of each subtest to the overall test variance:

$$R = 1 - \frac{\sum \sigma_j^2 (1 - \alpha_j)}{\sigma_X^2}$$

where σ_j^2 is the variance of stratum (or subtest) j, α_j is the reliability of stratum j, and σ_X^2 is the variance of the test as a whole. Reliabilities were computed by dividing the test into three strata: multiple-choice items, Composition scores, and Revising and Editing scores. This division was chosen based on both conceptual grounds and test-component variance-covariance matrices.

Tables 13-1 through 13-6 provide reliability information on the writing test for the total student

population and for students in each gender group and the ethnicity groups of White and Black, Hispanic, Asian, and Indian. Other groups such as LEP, IEP, and Economically Disadvantaged were also included for reliability estimation. The contents of the table include total number of points (K), number of students tested (N), mean points received, standard deviation (SD), mean P-Value, reliability, traditional standard error of measurement, and item type.

Reliabilities were fairly consistent across groups, though they trended lower for girls and white non-Hispanic students. The grade 5 reliabilities trended somewhat lower than the reliabilities for grades 8 and 10.

Table 13-1. GRADE 5 WRITING

Overall

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	131463	68.57	12.99	0.75	0.75	6.44	MC,CR

By gender

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Male	100	66990	66.16	13.10	0.73	0.76	6.48	MC,CR
Female	100	64278	71.11	12.37	0.77	0.73	6.43	MC,CR

By ethnicity

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
White non-Hispanic	100	98355	70.32	12.42	0.77	0.73	6.43	MC,CR
Black/African American non-Hispanic	100	20204	62.17	12.66	0.65	0.73	6.59	MC,CR
Latino/Hispanic	100	8341	62.29	13.10	0.65	0.76	6.40	MC,CR
Asian or Pacific Islander	100	3344	73.04	13.03	0.80	0.76	6.42	MC,CR
American Indian or Alaskan Native	100	168	66.30	13.50	0.72	0.75	6.71	MC,CR
Multi-Racial/Ethnic	100	791	65.19	13.12	0.70	0.76	6.37	MC,CR

LEP

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	2793	58.04	12.99	0.59	0.74	6.62	MC,CR

IEP

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	20293	57.22	13.58	0.59	0.77	6.51	MC,CR

ECO

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	47416	63.14	12.81	0.67	0.74	6.50	MC,CR

Table 13-2. GRADE 8 WRITING

Overall

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	142960	71.63	14.13	0.74	0.81	6.15	MC,CR

By gender

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Male	100	73147	68.71	14.32	0.72	0.82	6.12	MC,CR
Female	100	69512	74.76	13.22	0.76	0.78	6.19	MC,CR

By ethnicity

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
White non-Hispanic	100	109152	73.75	13.15	0.77	0.78	6.10	MC,CR
Black/African American non-Hispanic	100	21503	63.53	14.39	0.63	0.80	6.35	MC,CR
Latino/Hispanic	100	7999	63.54	14.90	0.62	0.82	6.25	MC,CR
Asian or Pacific Islander	100	3123	76.64	13.80	0.80	0.79	6.26	MC,CR
American Indian or Alaskan Native	100	246	68.76	14.14	0.71	0.81	6.13	MC,CR
Multi-Racial/Ethnic	100	551	67.28	15.30	0.68	0.82	6.46	MC,CR

LEP

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	2054	57.27	14.71	0.54	0.81	6.37	MC,CR

IEP

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	20809	57.43	14.64	0.56	0.81	6.36	MC,CR

ECO

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	46569	65.07	14.38	0.65	0.81	6.23	MC,CR

Table 13-3. GRADE 11 WRITING

Overall

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	132210	71.84	13.24	0.77	0.79	6.11	MC,CR

By gender

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Male	100	66480	69.77	13.72	0.75	0.80	6.17	MC,CR
Female	100	65412	73.97	12.36	0.78	0.76	6.03	MC,CR

By ethnicity

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
White non-Hispanic	100	107157	73.32	12.50	0.79	0.77	6.02	MC,CR
Black/African American non-Hispanic	100	15454	63.99	13.85	0.65	0.78	6.44	MC,CR
Latino/Hispanic	100	5203	64.15	14.11	0.66	0.80	6.35	MC,CR
Asian or Pacific Islander	100	3146	74.70	13.90	0.79	0.79	6.37	MC,CR
American Indian or Alaskan Native	100	215	68.05	13.90	0.72	0.78	6.49	MC,CR
Multi-Racial/Ethnic	100	643	66.92	14.01	0.72	0.80	6.34	MC,CR

LEP

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	1171	57.14	14.40	0.56	0.78	6.79	MC,CR

IEP

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	15961	57.55	14.43	0.57	0.79	6.55	MC,CR

ECO

Strand	K	N	Mean	SD	Mean P-Value	Reliability	SEM	Item Types In Strand
Overall	100	30034	65.31	13.93	0.68	0.80	6.30	MC,CR

Chapter Fourteen: Performance Level Validation Report

PURPOSE AND OBJECTIVES

The stated objective of the meeting was to set cutpoints for Writing in grades 5, 8, and 11. It was further stated that the results from this meeting would be presented to the State Board for review and possible adoption for application to student data in Spring 2006.

BODY OF WORK

DRC conducted the standard setting using the Body of Work method. This method is an examinee-centered one in which panelists are asked to evaluate a student's entire response set, both multiple-choice and open-ended prompt responses, in order to place that examinee into one of Below Basic, Basic, Proficient, or Advanced categories. The evaluation materials used by the panelists consisted of a sample of examinees with complete data sets in the 2006 administration who were chosen randomly from each score point between 27 and 100. Panelists were not informed of the examinee's original prompt scores, though they were 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 were presented with the entire sample of evaluation materials and asked to place each examinee into one of the four performance levels. In the pinpointing phase, DRC analysts determined, for each panelist, the scale-score cut points resulting from their full-sample classifications. Panelists were then presented with a collection of response sets consisting of the two examinees nearest each panelist's individual cut points, such that all panelists reviewed two examinee response sets for each cut point suggested in the rangefinding phase. All panelists reviewed the same collection of response sets – for example, if Panelist A proposed setting the cut point at point X and Panelist B proposed 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 were asked to make classification decisions by cut point; for example, for the response sets surrounding the Basic/Proficient cut point, panelists were asked to judge whether each student's response set fell into the Basic category or the Proficient one.

TRAINING

Training was conducted on the first morning of the meeting. Panelists were told that they were to:

- Be responsible for all secure materials
- Verify their individual placements for each round of judgments, and
- Participate in a discussion as a large group

Training materials included:

- Performance Level Descriptors (PLD)
- Sample student work
- Rating form including the student's responses to multiple-choice questions

Panelists were told that the process would include iterations (rounds) of individual judgments, small group discussions and large group discussions, and opportunities to revise judgments. In

addition, impacts would be presented (percent of students in each performance level) based on the panelists' judgments; impacts would also, when appropriate, be presented for other grades.

A copy of the agenda for the meeting is provided in Appendix P.

COMPOSITION OF PANEL

Twenty-nine panelists participated in the session. The demographics of the panelists are displayed in the following section. They include gender, role (e.g., teacher, educator/non-teacher, or other), region of residence in the state, and years of teaching experience. The following information is based on a self-report survey completed by the panelists; note that the region categories are self-report options on the demographic surveys and are not defined for participants.

DEMOGRAPHIC DISTRIBUTION

Figure 14.1

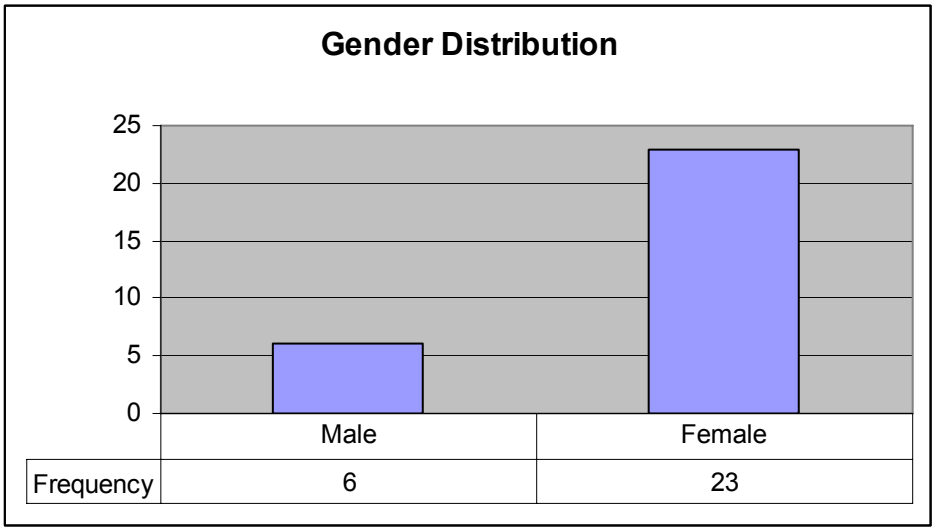


Figure 14.2

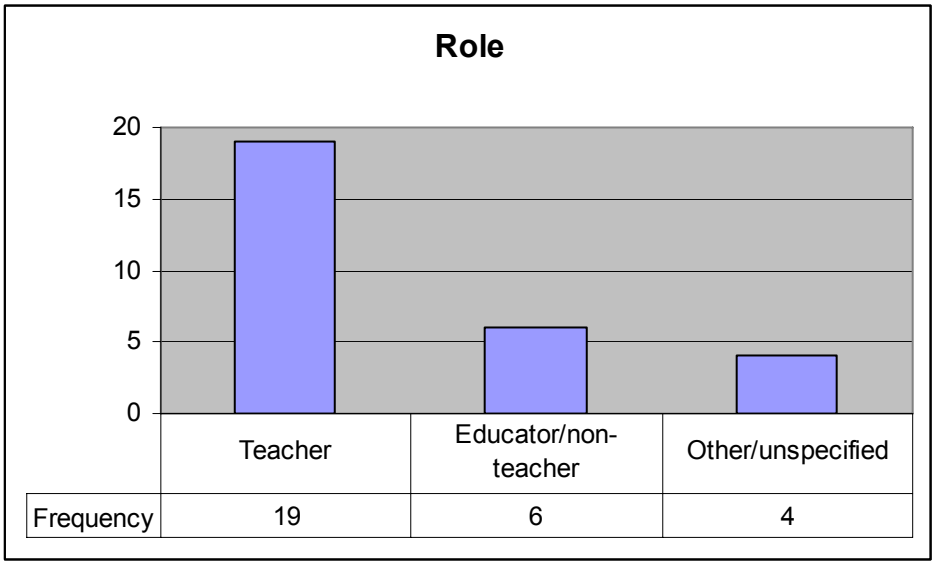


Figure 14.3

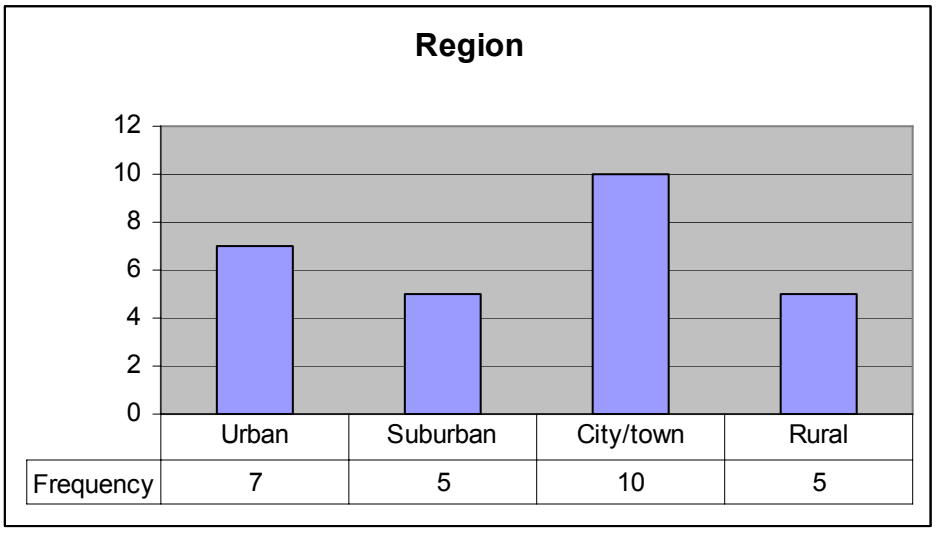


Figure 14.4

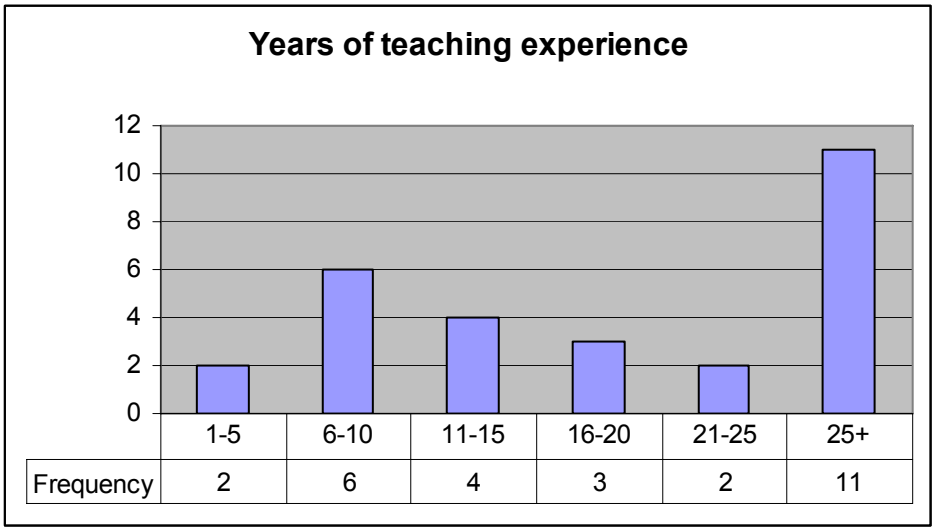
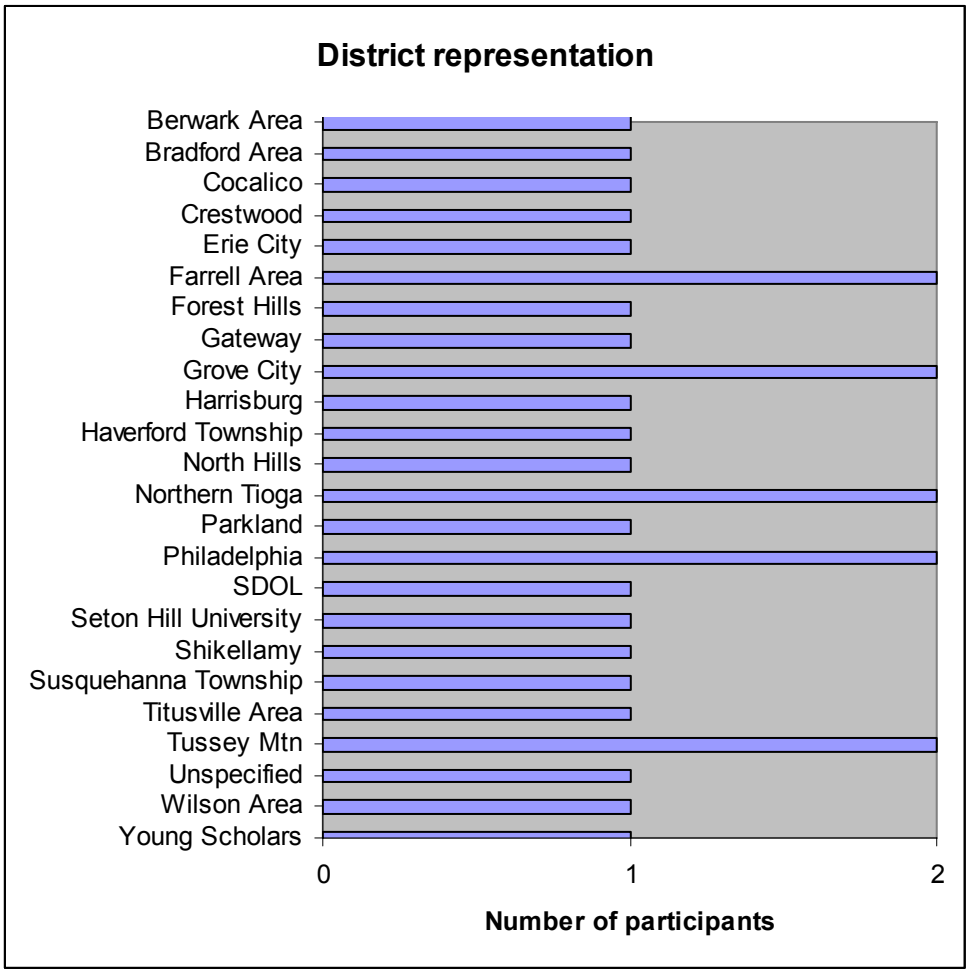


Figure 14.5



BREAKOUT OF PANELIST GROUPS

Approximately one-third of the participants reviewed grade 5, one-third reviewed Grade 8, and one-third reviewed Grade 11. Panelists first engaged in a rangefinding round, in which they were given one student response set (essay and multiple-choice responses) for each available score point after chance level. Panelists were asked to classify each paper into Below Basic, Basic, Proficient, and Advanced, and also to mark whether they felt each paper was high, medium, or low in its assigned category.

In the pinpointing phase, which took place on Day 2, DRC analysts determined, for each panelist, the scale-score cut points resulting from their full-sample classifications. Panelists were then presented with a collection of response sets consisting of the two examinees nearest each panelist's individual cut points, such that all panelists reviewed two examinee response sets for each cut point suggested in the rangefinding phase. All panelists reviewed the same collection of response sets – for example, if Panelist A proposed setting the cut point at point X and Panelist B proposed 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 were asked to make classification decisions by cut point; for example, for the response sets surrounding the Basic/Proficient cut point, panelists were asked to judge whether each student's response set fell into the Basic category or the Proficient one.

Results were shown to each group after both rangefinding and pinpointing. In addition, a checkpoint at the end of the final round, across all grades within subject, was added to the process. As part of this step, panelists were asked to assess whether they were confident in the resulting articulation of cutpoints across all grades.

CUTPOINTS AND STANDARD ERRORS

Cutpoints were derived using logistic regression, a statistical technique that 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 being proficient or above. 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 slop and intercept, respectively, of the logistic regression, and x is the score of interest. SPSS was used to estimate these parameters. After each round, the judges' binary decisions were used to estimate individual cut score estimates for each category in the logit metric. The medians of the individual estimates were then 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 was chosen as a representative sample. Note that the panelists were split into three groups that worked on grades 5, 8, or 11. Therefore, the sample size for each group was approximately 10.

It is important to note that the calculations were based on the standard error of the median. The standard error of the median, given a normal distribution or large samples, is approximately 25

percent larger than the standard error of the means. Thus, the standard error of the means was 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 grades, the standard errors were pooled across grades within subject and cutpoint.

Table 14-1 shows the standard errors in the logit (Rasch) metric for each cutpoint, by grade. Table 14-2 shows the same information, but in the scale score metric.

Table 14-1. Standard Errors Logit Metric

	Standard Errors – Logit Metric		
Grade	**BB/B**	**B/P**	**P/A**
5	.40	.38	.56
8	.42	.39	.44
11	.38	.34	.49
Pooled	.40	.37	.50

Table 14-2. Standard Errors Scale Score Metric

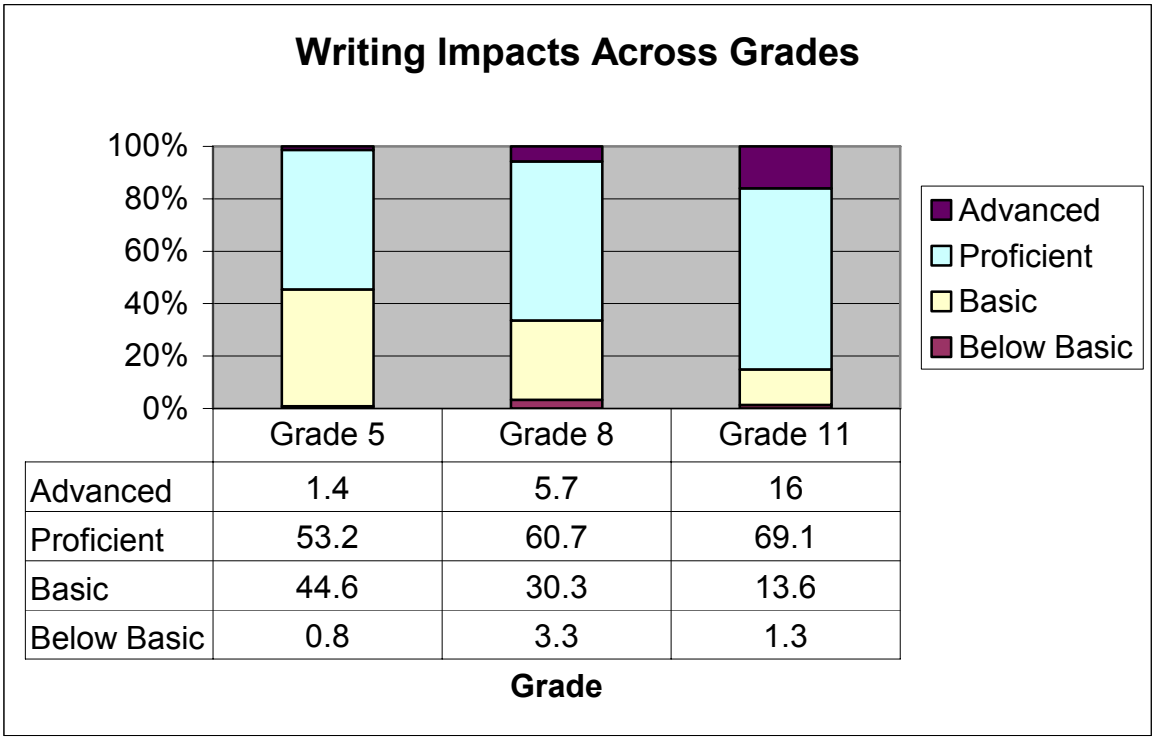
	Standard Errors – Scale Score		
Grade	**BB/B**	**B/P**	**P/A**
5	40	38	56
8	42	39	44
11	38	34	49
Pooled	40	37	50

BB- Below Basic
B- Basic
P- Proficient
A- Advanced

RESULTS

This section presents plots that show the recommendations taken to the State Board on June 29, 2006. The results shown below are based on the percentage of students in each performance level.

Figure 14.6



	Grade 5	Grade 8	Grade 11
Advanced	1.4	5.7	16
Proficient	53.2	60.7	69.1
Basic	44.6	30.3	13.6
Below Basic	0.8	3.3	1.3

SCALING AND TRANSFORMATIONS

Table 14-3 shows the linear equations used to convert student scores from the logit metric to the scale score metric and Table 14-4 shows the scale score cutpoints for each grade and subject.

Table 14-3. Conversion Equations

Grade	Conversion Equation
5	Y = 100X + 1071.44
8	Y = 100X + 1123.84
11	Y = 100X + 1244.30

Table 14-4. Scale Score Cutpoints

Performance Level	Grade 5	Grade 8	Grade 11
Advanced	1909 and up	1748 and up	1806 and up
Proficient	1236-1908	1236-1747	1236-1805
Basic	745-1235	914-1235	952-1235
Below Basic	744 and below	913 and below	951 and below

PANELIST EVALUATION SURVEY RESULTS

Summary results may be found in Appendix Q. The appendix contains question-by-question summary ratings that reflect the panelists' level of satisfaction with the method, materials, training, process, individual and group judgments and recommendations, facilities, food, and use of time. The majority of panelists reported being "confident" or "very confident" in their final judgments and "confident" or "very confident" in the ability of the process to produce reliable and valid results. Panelists also reported being satisfied with the training and finding the materials provided useful.

References

Achieve, Inc. (2005). Measuring Up 2005: A Report on Assessment Anchors and Tests in Reading and Mathematics for Pennsylvania. Washington, DC: Achieve, Inc.

AERA, APA, NCME (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education). (1999). *Standards for Educational and Psychological Tests*. Washington, DC: American Educational Research Association.

Allman, C. (2004). Test access: Making tests accessible for students with visual impairments – A guide for test publishers, test developers, and state assessment personnel (2nd edition). Louisville, KY: American Printing House for the Blind. Available from http://www.aph.org.

Bhola, D. L., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: methods and issues. *Educational Measurement: Issues and Practice, 22* (3), 21-29.

Cronbach, L. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16, 297-334.

Data Recognition Corporation. (2000). *Item Viewer and Authoring Network (IVAN): Informational Guide*. Maple Grove, MN: DRC.

Data Recognition Corporation. (2003-2006). *Fairnes*s *in Testing: Training Manual for Issues of Bias, Fairness, and Sensitivity*. Maple Grove, MN: DRC.

Data Recognition Corporation. (2004-2006). *Pennsylvania System of School Assessment (PSSA) Style Guide*. Maple Grove, MN: DRC.

Data Recognition Corporation. (2005, March). *Preliminary Technical Report for the PSSA 2005 Reading and Mathematics*. Maple Grove, MN: DRC.

Data Recognition Corporation. (2005, October). *Preliminary Technical Report for the PSSA 2006 Writing*. Maple Grove, MN: DRC.

Dorans, N., Schmitt, Al, & Bleistein, C. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, 29, 309-319.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Pennsylvania State Board of Education. (1999, January). *Chapter 4. Academic Standards and Assessment*. Harrisburg, PA: Pennsylvania State Board of Education. Retrieved November 8, 2004, from http://www.pde.state.pa.us. Also available from http://www.pacode.com/secure/data/022/Chapter4/s4.51.html.

Pennsylvania Department of Education. (2005, December). *2006 Accommodations Guidelines for Students with IEPs, Students with 504 Plans and English Language Learners*. Harrisburg, PA: PDE. Retrieved January 30, 2006, from http://www.pde.state.pa.us.

Pennsylvania Department of Education. (2005, December). 2005-2006 *Writing Assessment Handbook*. Harrisburg, PA: PDE. Retrieved January 30, 2006, from http://www.pde.state.pa.us.

Pennsylvania Department of Education. (2005). 2005-2006 *Writing Item and Scoring Sampler*. Harrisburg, PA: PDE. Posted separately by grade level. Retrieved September 14, 2005, from http://www.pde.state.pa.us.

Pennsylvania Department of Education. (2006, February). *PSSA 2006 Handbook for Assessment Coordinators and Administrators: Grades 5, 8, and 11 Writing*. Harrisburg, PA: Retrieved January 30, 2006, from http://www.pde.state.pa.us.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.

Thacker, A. A. & Dickinson, E. R. (2004). *Item Content and Difficulty Mapping by Form and Item Type for the 2001-2003 Pennsylvania System of School Assessment (PSSA)*. Alexandria, VA: Human Resources Research Organization.

Thacker, A. A., Dickinson, E. R. & Koger, M. E. (2004). *Relationships Among the Pennsylvania System of School Assessment (PSSA) and Other Commonly Administered Assessments.* Alexandria, VA: Human Resources Research Organization.

Thompson, S., Johnstone, C. J. & Thurlow, M. L. (2002). *Universal Design Applied to Large Scale Assessments* (Synthesis Report 44), Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Webb, N. L. (1999). *Research Monograph No. 18: Alignment of Science and Mathematics Standards and Assessments in Four States.* Madison, WI: National Institute for Science Education.

WINSTEPS (2000). *WINSTEPS® Rasch Measurement.* Copyright John M. Linacre.

Wright, B. & Masters, G. (1982). *Rating Scale Analysis*. Chicago: MESA Press.

Appendix A:

PSSA Writing Scoring Guidelines

PSSA CONVENTIONS SCORING GUIDELINE

4

- Thorough control of sentence formation.
- Few errors, if any, are present in grammar, usage, spelling, and punctuation, but the errors that are present do not interfere with meaning.

3

- Adequate control of sentence formation.
- Some errors may be present in grammar, usage, spelling, and punctuation, but few, if any, of the errors that are present may interfere with meaning.

2

- Limited and/or inconsistent control of sentence formation. Some sentences may be awkward or fragmented
- Many errors may be present in grammar, usage, spelling, and punctuation, and some of those errors may interfere with meaning.

1

- Minimal control of sentence formation. Many sentences are awkward and fragmented.
- Many errors may be present in grammar, usage, spelling, and punctuation, and many of those errors may interfere with meaning.

PSSA INFORMATIONAL SCORING GUIDELINE

4

FOCUS	Sharp, distinct controlling point made about a single topic with evident awareness of task and audience.
CONTENT DEVELOPMENT	Substantial, relevant, and illustrative content that demonstrates a clear understanding of the purpose. Thorough elaboration with effectively presented information consistently supported with well-chosen details.
ORGANIZATION	Effective organizational strategies and structures, such as logical order and transitions, which develop a controlling idea.
STYLE	Precise control of language, stylistic techniques, and sentence structures that creates a consistent and effective tone.

3

FOCUS	Clear controlling point made about a single topic with general awareness of task and audience.
CONTENT DEVELOPMENT	Adequate, specific, and/or illustrative content that demonstrates an understanding of the purpose. Sufficient elaboration with clearly presented information supported with well-chosen details.
ORGANIZATION	Organizational strategies and structures, such as logical order and transitions, which develop a controlling idea.
STYLE	Appropriate control of language, stylistic techniques, and sentence structures that creates a consistent tone.

2

FOCUS	Vague evidence of a controlling point made about a single topic with an inconsistent awareness of task and audience.
CONTENT DEVELOPMENT	Inadequate, vague content that demonstrates a weak understanding of the purpose. Underdeveloped and/or repetitive elaboration with inconsistently supported information. May be an extended list.
ORGANIZATION	Inconsistent organizational strategies and structures, such as logical order and transitions, which ineffectively develop a controlling idea.
STYLE	Limited control of language and sentence structures that creates interference with tone.

1

FOCUS	Little or no evidence of a controlling point made about a single topic with a minimal awareness of task and audience.
CONTENT DEVELOPMENT	Minimal evidence of content that demonstrates a lack of understanding of the purpose. Superficial, undeveloped writing with little or no support. May be a bare list.
ORGANIZATION	Little or no evidence of organizational strategies and structures, such as logical order and transitions, which inadequately develop a controlling idea.
STYLE	Minimal control of language and sentence structures that creates an inconsistent tone.

PSSA NARRATIVE SCORING GUIDELINE

4

FOCUS	Sharp, distinct controlling point or theme with evident awareness of the narrative.
CONTENT DEVELOPMENT	Strong story line with illustrative details that addresses a complex idea or examines a complex experience. Thoroughly elaborated narrative sequence that employs narrative elements as appropriate.
ORGANIZATION	Skillful narrative pattern with clear and consistent sequencing of events, employing a beginning, a middle, and an end. Minor interruptions to the sequence may occur.
STYLE	Precise control of language, literary devices, and sentence structures that creates a consistent and effective point of view and tone.

3

FOCUS	Clear controlling point or theme with general awareness of the narrative.
CONTENT DEVELOPMENT	Story line with details that addresses an idea or examines an experience. Sufficiently elaborated narrative sequence that employs narrative elements as appropriate.
ORGANIZATION	Narrative pattern with generally consistent sequencing of events, employing a beginning, a middle, and an end. Interruptions to the sequence may occur.
STYLE	Appropriate control of language, literary devices, and sentence structures that creates a consistent point of view and tone.

2

FOCUS	Vague evidence of a controlling point or theme with inconsistent awareness of the narrative.
CONTENT DEVELOPMENT	Inconsistent story line that inadequately addresses an idea or examines an experience. Insufficiently elaborated narrative sequence that may employ narrative elements.
ORGANIZATION	Narrative pattern with generally inconsistent sequencing of events that may employ a beginning, a middle, and an end. Interruptions to the sequence may interfere with meaning.
STYLE	Limited control of language and sentence structures that creates interference with point of view and tone.

1

FOCUS	Little or no evidence of a controlling point or theme with minimal awareness of the narrative.
CONTENT DEVELOPMENT	Insufficient story line that minimally addresses an idea or examines an experience. Unelaborated narrative that may employ narrative elements.
ORGANIZATION	Narrative pattern with little or no sequencing of events. Interruptions to the sequence interfere with meaning.
STYLE	Minimal control of language and sentence structures that creates an inconsistent point of view and tone.

PSSA PERSUASIVE SCORING GUIDELINE

4

FOCUS	Sharp, distinct controlling point presented as a position and made convincing through a clear, thoughtful, and substantiated argument with evident awareness of task and audience.
CONTENT DEVELOPMENT	Substantial, relevant, and illustrative content that demonstrates a clear understanding of the purpose. Thoroughly elaborated argument that includes a clear position consistently supported with precise and relevant evidence. Rhetorical (persuasive) strategies are evident.
ORGANIZATION	Effective organizational strategies and structures, such as logical order and transitions, to develop a position supported with a purposeful presentation of content.
STYLE	Precise control of language, stylistic techniques, and sentence structures that creates a consistent and effective tone.

3

FOCUS	Clear controlling point presented as a position and made convincing through a credible and substantiated argument with general awareness of task and audience.
CONTENT DEVELOPMENT	Adequate, specific and/or illustrative content that demonstrates an understanding of the purpose. Sufficiently elaborated argument that includes a clear position supported with some relevant evidence. Rhetorical (persuasive) strategies may be evident.
ORGANIZATION	Organizational strategies and structures, such as logical order and transitions, to develop a position supported with sufficient presentation of content.
STYLE	Appropriate control of language, stylistic techniques, and sentence structures that creates a consistent tone.

2

FOCUS	Vague evidence of a controlling point presented as a position that may lack a credible and/or substantiated argument with an inconsistent awareness of task and audience.
CONTENT DEVELOPMENT	Inadequate, vague content that demonstrates a weak understanding of the purpose. Insufficiently elaborated argument that includes an underdeveloped position supported with little evidence.
ORGANIZATION	Inconsistent organizational strategies and structures, such as logical order and transitions, to develop a position with inadequate presentation of content.
STYLE	Limited control of language and sentence structures that creates interference with tone.

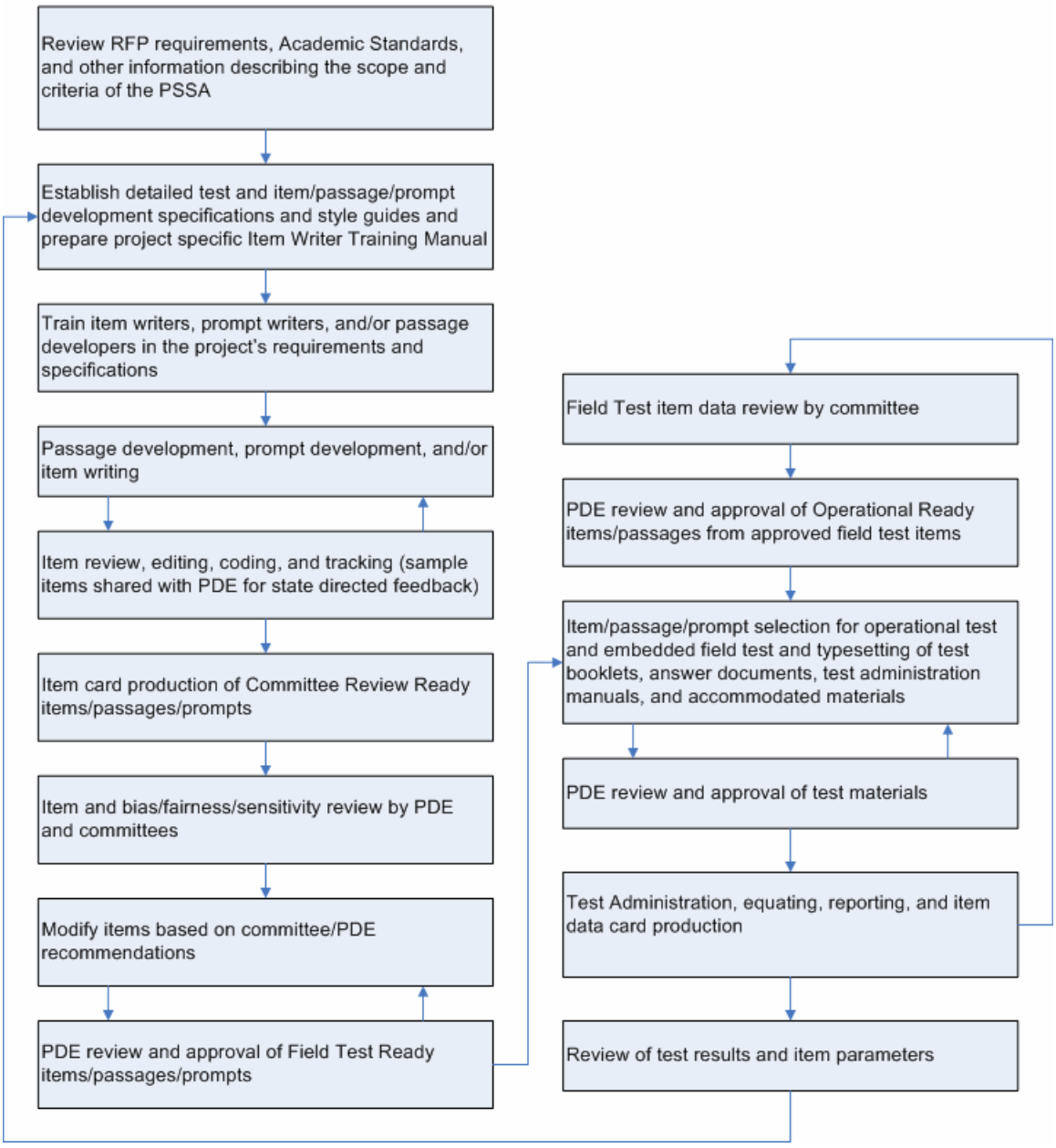
1

FOCUS	Little or no evidence of a controlling point presented as a position that lacks a credible and/or substantiated argument with minimal awareness of task and audience.
CONTENT DEVELOPMENT	Minimal evidence of content that demonstrates a lack of understanding of the purpose. Unelaborated argument that includes an undeveloped position supported with minimal or no evidence.
ORGANIZATION	Little or no evidence of organizational strategies and structures, such as logical order and transitions, to develop a position with insufficient presentation of content.
STYLE	Minimal control of language and sentence structures that creates an inconsistent tone.

Appendix B:

Item and Test Development Processes

DRC Item and Test Development Process
Writing



Appendix C:

Item Review Form

Item Review Form

Grades: 5 8 11 **Content:** Math Reading Writing

	Content Alignment		Rigor Level Alignment			Technical Design			STATUS
Unique ID number	Anchor (A.1.3, B.2.1, etc.)	Content Limits (A.1.3.2, B.2.1.2)	Grade	Depth of Knowledge	Source of Challenge or Language Demand	Correct Answer	Distracters	Graphics	Acceptance Status
									A=Accept E=Edit

Comments/Recommendations

Appendix D:

Item Review Cards and IVAN Card

Client:	50	Item ID:
		(for internal use only)

Item Writing Form - Pennsylvania

Item Writer Information		Stimulus	
Item Writer Name		Stimulus Used	
Item Writer #		Stimulus ID	
Email Address		Stimulus Title	
Submission Date		Stimulus Type	
Item Information		Delivery	Atch ☐ Fax ☐ Mail ☐
Item Type		Title	
Grade		Author	
Subject		Publisher	
Goal 1		Date Published	
Goal 2		Source Page	
Goal 3		URL	
Goal 4		Permission Needed	
Taxonomy Level			
Depth of Knowledge		Passage Title #1	
Difficulty		Temp Passage ID #1	
Focus			
Graphics		Passage Title #2	
Calculator		Temp Passage ID #2	
Points			

Comment	

Prompt / Stem

Answer Options	
Key:	
Option A. Rationale:	
Option B. Rationale:	
Option C. Rationale:	
Option D. Rationale:	

Rubric

Client:	50	Item ID: *(for internal use only)*

Passage Writing Form - Pennsylvania

Passage Writer Information		Reference	
Passage Writer Name		**Delivery**	Atch ☐ Fax ☐ Mail ☐
Passage Writer #		**Title**	
Email Address		**Publisher**	
Submission Date		**Date Published**	
Item Information		**Source Page**	
Subject		**URL**	
Grade		**Permission Needed**	
Type			
Category		**Paired Passage**	
Temp. Passage ID			
Passage Title			
Passage Author			

Comment	

Passage

IVAN Item Card

DATA RECOGNITION DRC CORPORATION

Released: **Item Status:**

Item Name	Item Type	Key	Grade	Subject	Report Category	Asmt Anchor	Sub - Anchor	Eligible Content	Content Difficulty	DRP	Item Calculator

Depth of Knowledge:

Administration

Form Grade	Form Subject	Form Name	Sequence	Form Type	Month	Year	Report Category	Asmt Anchor	Sub-Anchor	Eligible Content	Day	Session	Calculator

Statistics Detail

Label	P-Value	Pt. Bis. Corr.
0		
1		
2		
3		
4		
Omits		
Mean		

Label	Value
N	
Outfit t	
Logit	

DIF Analysis	Value
White/Black	
Male/Female	

Notes:

Appendix E:

2006 Uncommon Grade 5 Multiple Choice Statistics for Writing

Item Detail			Proportions						Point Biserials					Rasch Statistics		
Item ID	Answer Key	N	P-Value	A	B	C	D	Other	Item Total Corr	A	B	C	D	Logit	SE	Outfit t
02307	B	13217	0.5239	0.1282	0.5239	0.2147	0.1313	0.0017	0.2978	-0.3051	0.2978	-0.2110	-0.2953	1.8936	0.0316	9.90
02308	B	13217	0.5348	0.0157	0.5348	0.4142	0.0342	0.0010	0.0983	-0.2258	0.0983	-0.0616	-0.1309	1.8654	0.0316	9.90
02309	A	13217	0.8609	0.8609	0.0369	0.0552	0.0455	0.0014	0.3657	0.3657	-0.2670	-0.2600	-0.2114	-0.8473	0.0428	9.90
02310	A	13217	0.9396	0.9396	0.0146	0.0263	0.0160	0.0034	0.3648	0.3648	-0.2027	-0.2547	-0.2375	-2.2028	0.0626	5.30
02311	B	13217	0.5911	0.1257	0.5911	0.1528	0.1281	0.0023	0.3256	-0.2766	0.3256	-0.2696	-0.2671	1.4222	0.0318	9.90
02312	A	13217	0.8168	0.8168	0.0666	0.0891	0.0262	0.0014	0.4134	0.4134	-0.3093	-0.2915	-0.2935	-0.4245	0.0391	9.90
02313	D	13217	0.8320	0.0899	0.0384	0.0383	0.8320	0.0014	0.3821	-0.2658	-0.2711	-0.2759	0.3821	-0.4540	0.0393	9.90
02314	A	13217	0.6339	0.6339	0.2038	0.0950	0.0642	0.0033	0.3612	0.3612	-0.2309	-0.3344	-0.3966	1.0923	0.0323	9.90
02315	B	13182	0.7516	0.1026	0.7516	0.0804	0.0636	0.0018	0.3942	-0.2838	0.3942	-0.2912	-0.3120	0.2828	0.0353	9.90
02316	B	13182	0.7509	0.1523	0.7509	0.0495	0.0463	0.0009	0.3795	-0.3095	0.3795	-0.2672	-0.2667	0.2841	0.0352	9.90
02317	A	13182	0.7180	0.7180	0.1287	0.0786	0.0734	0.0012	0.4134	0.4134	-0.2286	-0.3777	-0.3966	0.4808	0.0344	9.90
02318	C	13182	0.7987	0.0359	0.1311	0.7987	0.0320	0.0023	0.3104	-0.2378	-0.2180	0.3104	-0.2412	-0.1892	0.0379	9.90
02319	A	13182	0.7224	0.7224	0.0299	0.1878	0.0583	0.0016	0.3674	0.3674	-0.3583	-0.2664	-0.2991	0.4449	0.0345	9.90
02320	B	13182	0.4583	0.1714	0.4583	0.3046	0.0646	0.0011	0.3086	-0.2168	0.3086	-0.3180	-0.3512	2.3950	0.0318	9.90
02321	B	13182	0.7112	0.0450	0.7112	0.1364	0.1050	0.0024	0.3795	-0.2805	0.3795	-0.2964	-0.3063	0.5659	0.0341	9.90
02322	C	13182	0.6708	0.1268	0.1158	0.6708	0.0827	0.0038	0.3886	-0.3277	-0.2920	0.3886	-0.3133	0.8714	0.0331	9.90
02323	A	13131	0.8823	0.8823	0.0702	0.0241	0.0227	0.0007	0.3543	0.3543	-0.2219	-0.2480	-0.2784	-1.0229	0.0453	9.90
02324	A	13131	0.6900	0.6900	0.2812	0.0196	0.0089	0.0002	0.3654	0.3654	-0.3283	-0.3176	-0.2058	0.7832	0.0335	9.90
02325	A	13131	0.8025	0.8025	0.0730	0.0544	0.0695	0.0008	0.4223	0.4223	-0.3018	-0.3035	-0.3125	-0.2192	0.0383	9.90
02326	C	13131	0.8333	0.0354	0.1026	0.8333	0.0281	0.0006	0.3387	-0.2857	-0.2469	0.3387	-0.1732	-0.4376	0.0399	9.90
02327	A	13131	0.9319	0.9319	0.0241	0.0242	0.0192	0.0005	0.3749	0.3749	-0.2464	-0.2345	-0.2401	-1.8640	0.0569	5.90
02328	A	13131	0.5543	0.5543	0.1641	0.1330	0.1472	0.0014	0.2765	0.2765	-0.2797	-0.2387	-0.2016	1.7843	0.0318	9.90
02329	B	13131	0.7817	0.1642	0.7817	0.0420	0.0106	0.0014	0.3873	-0.3200	0.3873	-0.3052	-0.1994	0.0207	0.0368	9.90
02330	B	13131	0.6609	0.0752	0.6609	0.1269	0.1337	0.0034	0.3745	-0.2177	0.3745	-0.3825	-0.2820	0.9947	0.0329	9.90
02331	B	13157	0.7100	0.0216	0.7100	0.0311	0.2364	0.0009	0.2971	-0.2758	0.2971	-0.2870	-0.2218	0.5819	0.0338	9.90
02332	A	13157	0.8070	0.8070	0.0901	0.0489	0.0530	0.0010	0.4360	0.4360	-0.3071	-0.3341	-0.3209	-0.2999	0.0383	9.90
02333	C	13157	0.5494	0.2656	0.1129	0.5494	0.0703	0.0017	0.2460	-0.1601	-0.2700	0.2460	-0.2714	1.7605	0.0318	9.90
02334	C	13157	0.7210	0.0624	0.0777	0.7210	0.1374	0.0015	0.3948	-0.3246	-0.3153	0.3948	-0.2862	0.4913	0.0341	9.90
02335	B	13157	0.9190	0.0289	0.9190	0.0229	0.0286	0.0007	0.3669	-0.2229	0.3669	-0.2755	-0.2166	-1.6605	0.0533	8.10
02336	B	13157	0.7879	0.0720	0.7879	0.1213	0.0180	0.0008	0.4585	-0.3417	0.4585	-0.3865	-0.2455	-0.1071	0.0370	9.90
02337	B	13157	0.8667	0.0336	0.8667	0.0383	0.0605	0.0009	0.4346	-0.2642	0.4346	-0.3048	-0.3206	-0.8976	0.0435	8.10
02338	C	13157	0.7200	0.1642	0.0522	0.7200	0.0608	0.0027	0.3071	-0.2441	-0.2073	0.3071	-0.2349	0.5346	0.0340	9.90
02339	A	13105	0.5042	0.5042	0.1776	0.0529	0.2641	0.0011	0.2068	0.2068	-0.1871	-0.2058	-0.1770	2.1788	0.0313	9.90
02340	B	13105	0.7320	0.0489	0.7320	0.0721	0.1462	0.0008	0.3797	-0.2716	0.3797	-0.3278	-0.2754	0.4717	0.0344	9.90
02341	C	13105	0.8931	0.0482	0.0291	0.8931	0.0287	0.0009	0.4262	-0.2982	-0.2877	0.4262	-0.2667	-1.2347	0.0485	2.40
02342	D	13105	0.4975	0.0808	0.3695	0.0498	0.4975	0.0024	0.1067	-0.2451	-0.0211	-0.2031	0.1067	2.1828	0.0313	9.90
02343	A	13105	0.8052	0.8052	0.0547	0.0460	0.0929	0.0012	0.4273	0.4273	-0.3161	-0.3408	-0.2820	-0.1808	0.0381	9.90
02344	D	13105	0.5844	0.2360	0.0969	0.0812	0.5844	0.0015	0.3775	-0.2677	-0.4341	-0.3349	0.3775	1.5530	0.0316	9.90
02345	D	13105	0.5316	0.1373	0.1176	0.2115	0.5316	0.0020	0.3939	-0.3326	-0.4457	-0.3145	0.3939	1.9487	0.0313	9.90
02346	D	13105	0.5740	0.0954	0.1309	0.1950	0.5740	0.0047	0.3120	-0.2733	-0.3023	-0.2375	0.3120	1.6322	0.0315	9.90
02347	C	13166	0.8090	0.0575	0.0777	0.8090	0.0545	0.0013	0.3273	-0.2207	-0.2300	0.3273	-0.2402	-0.2359	0.0378	9.90
02348	C	13166	0.8626	0.0347	0.0719	0.8626	0.0300	0.0008	0.4043	-0.2460	-0.3108	0.4043	-0.2599	-0.8595	0.0430	9.60
02349	B	13166	0.6298	0.1798	0.6298	0.0963	0.0925	0.0016	0.2734	-0.1732	0.2734	-0.2832	-0.2395	1.2326	0.0320	9.90
02350	C	13166	0.4709	0.1374	0.3395	0.4709	0.0498	0.0024	0.2458	-0.2852	-0.1739	0.2458	-0.3208	2.2904	0.0315	9.90
02351	B	13166	0.8337	0.0285	0.8337	0.0899	0.0463	0.0016	0.3587	-0.2744	0.3587	-0.2396	-0.2572	-0.4832	0.0396	9.90

Item Detail			Proportions						Point Biserials					Rasch Statistics		
Item ID	Answer Key	N	P-Value	A	B	C	D	Other	Item Total Corr	A	B	C	D	Logit	SE	Outfit t
02352	C	13166	0.8188	0.0678	0.0536	0.8188	0.0580	0.0018	0.4501	-0.3163	-0.3293	0.4501	-0.3262	-0.3447	0.0386	8.70
02353	A	13166	0.5080	0.5080	0.0768	0.1581	0.2547	0.0024	0.2510	0.2510	-0.1767	-0.2356	-0.2395	2.0605	0.0314	9.90
02354	D	13166	0.6667	0.1589	0.0894	0.0795	0.6667	0.0055	0.3149	-0.2213	-0.2673	-0.2820	0.3149	0.9661	0.0325	9.90
02355	B	13095	0.7490	0.1033	0.7490	0.0430	0.1035	0.0012	0.3978	-0.3378	0.3978	-0.2100	-0.3100	0.3156	0.0352	9.90
02356	B	13095	0.8276	0.0605	0.8276	0.0328	0.0784	0.0008	0.4749	-0.3385	0.4749	-0.2920	-0.3790	-0.4529	0.0397	9.50
02357	B	13095	0.9269	0.0342	0.9269	0.0216	0.0163	0.0009	0.3913	-0.2578	0.3913	-0.2585	-0.2428	-1.8534	0.0567	4.80
02358	D	13095	0.7591	0.1155	0.0364	0.0871	0.7591	0.0018	0.3436	-0.3145	-0.1314	-0.2521	0.3436	0.1873	0.0358	9.90
02359	C	13095	0.5859	0.1347	0.1436	0.5859	0.1337	0.0020	0.2901	-0.2048	-0.2924	0.2901	-0.2202	1.5921	0.0322	9.90
02360	C	13095	0.8344	0.0464	0.0284	0.8344	0.0889	0.0020	0.4313	-0.2642	-0.3375	0.4313	-0.3255	-0.5769	0.0407	9.90
02361	D	13095	0.7697	0.1231	0.0520	0.0526	0.7697	0.0026	0.4018	-0.2345	-0.3848	-0.3319	0.4018	0.0693	0.0364	9.90
02362	A	13095	0.8859	0.8859	0.0409	0.0233	0.0457	0.0043	0.4170	0.4170	-0.3033	-0.2647	-0.2774	-1.1674	0.0467	9.30
02363	C	13150	0.7625	0.0297	0.1262	0.7625	0.0805	0.0011	0.2615	-0.2376	-0.1941	0.2615	-0.1598	0.2609	0.0350	9.90
02364	A	13150	0.8896	0.8896	0.0430	0.0234	0.0432	0.0008	0.3813	0.3813	-0.2603	-0.2409	-0.2565	-1.2182	0.0474	4.30
02365	B	13150	0.8116	0.0865	0.8116	0.0468	0.0536	0.0014	0.3028	-0.2527	0.3028	-0.2024	-0.1649	-0.1991	0.0377	9.90
02366	B	13150	0.7833	0.1367	0.7833	0.0535	0.0246	0.0019	0.3122	-0.3118	0.3122	-0.1230	-0.1505	-0.0121	0.0365	9.90
02367	B	13150	0.7778	0.0659	0.7778	0.0859	0.0690	0.0014	0.3201	-0.2951	0.3201	-0.2384	-0.1537	0.0504	0.0361	9.90
02368	C	13150	0.7763	0.0716	0.1013	0.7763	0.0490	0.0018	0.4195	-0.2639	-0.3656	0.4195	-0.2744	-0.0040	0.0364	9.90
02369	C	13150	0.4132	0.2798	0.2424	0.4132	0.0611	0.0036	0.3244	-0.2916	-0.3330	0.3244	-0.3580	2.7164	0.0321	9.90
02370	D	13150	0.4876	0.1724	0.2039	0.1319	0.4876	0.0043	0.3357	-0.2870	-0.2873	-0.3741	0.3357	2.1810	0.0315	9.90
02371	C	13165	0.7775	0.1711	0.0235	0.7775	0.0270	0.0009	0.4258	-0.3639	-0.3068	0.4258	-0.2534	0.0102	0.0362	9.90
02372	D	13165	0.8223	0.0367	0.1044	0.0359	0.8223	0.0008	0.4071	-0.2356	-0.3434	-0.2443	0.4071	-0.3477	0.0384	9.90
02373	D	13165	0.7204	0.1337	0.0986	0.0461	0.7204	0.0012	0.3110	-0.2050	-0.2729	-0.2361	0.3110	0.5195	0.0340	9.90
02374	B	13165	0.8322	0.0652	0.8322	0.0304	0.0705	0.0017	0.3273	-0.3074	0.3273	-0.2532	-0.1280	-0.4977	0.0395	9.90
02375	D	13165	0.8836	0.0248	0.0555	0.0339	0.8836	0.0022	0.3707	-0.2673	-0.2302	-0.2614	0.3707	-1.0578	0.0448	7.70
02376	A	13165	0.5586	0.5586	0.1717	0.1612	0.1067	0.0017	0.2792	0.2792	-0.3178	-0.1299	-0.2844	1.7086	0.0319	9.90
02377	B	13165	0.7702	0.1433	0.7702	0.0400	0.0443	0.0023	0.3763	-0.2985	0.3763	-0.2821	-0.2497	0.0885	0.0358	9.90
02378	D	13165	0.8044	0.0402	0.0232	0.1275	0.8044	0.0048	0.3693	-0.2746	-0.2908	-0.2672	0.3693	-0.2071	0.0375	9.90
02379	A	13095	0.4001	0.4001	0.1940	0.2374	0.1676	0.0008	0.1136	0.1136	-0.0599	-0.1118	-0.1636	2.8044	0.0322	9.90
02380	B	13095	0.9346	0.0501	0.9346	0.0068	0.0079	0.0006	0.3707	-0.3277	0.3707	-0.1410	-0.1669	-1.9522	0.0585	2.40
02381	D	13095	0.7512	0.1399	0.0670	0.0406	0.7512	0.0012	0.3546	-0.2964	-0.2306	-0.2427	0.3546	0.2970	0.0351	9.90
02382	A	13095	0.6729	0.6729	0.1047	0.1103	0.1106	0.0015	0.3970	0.3970	-0.3765	-0.2764	-0.2972	0.9175	0.0329	9.90
02383	D	13095	0.7386	0.0385	0.0563	0.1651	0.7386	0.0015	0.2824	-0.2262	-0.2206	-0.2043	0.2824	0.4686	0.0344	9.90
02384	C	13095	0.7467	0.0799	0.1053	0.7467	0.0671	0.0010	0.4760	-0.3306	-0.3821	0.4760	-0.3756	0.2920	0.0351	9.90
02385	C	13095	0.6174	0.1163	0.1859	0.6174	0.0779	0.0024	0.3141	-0.2588	-0.2797	0.3141	-0.2243	1.2870	0.0321	9.90
02386	A	13095	0.3741	0.3741	0.0329	0.2471	0.3391	0.0067	0.1844	0.1844	-0.3209	-0.2099	-0.1385	3.0028	0.0327	9.90

Appendix F:

2006 Uncommon Grade 8 Multiple Choice Statistics for Writing

Item Detail			Proportions						Point Biserials					Rasch Statistics		
Item ID	Answer Key	N	P-Value	A	B	C	D	Other	Item Total Corr	A	B	C	D	Logit	SE	Outfit t
02387	C	14336	0.7999	0.0657	0.1221	0.7999	0.0116	0.0007	0.2839	-0.2367	-0.1995	0.2839	-0.1774	-0.2984	0.0368	9.90
02388	A	14336	0.7424	0.7424	0.0610	0.0375	0.1586	0.0005	0.3397	0.3397	-0.2816	-0.2165	-0.2667	0.2504	0.0340	9.90
02389	D	14336	0.5209	0.1312	0.2526	0.0940	0.5209	0.0013	0.3588	-0.3099	-0.3112	-0.3824	0.3588	1.8894	0.0302	9.90
02390	D	14336	0.6061	0.1800	0.1479	0.0646	0.6061	0.0014	0.2992	-0.2715	-0.2671	-0.1771	0.2992	1.2817	0.0308	9.90
02391	A	14336	0.8907	0.8907	0.0634	0.0208	0.0234	0.0017	0.3945	0.3945	-0.3064	-0.2594	-0.2047	-1.4929	0.0470	5.10
02392	C	14336	0.7824	0.1119	0.0657	0.7824	0.0389	0.0011	0.4503	-0.3530	-0.3465	0.4503	-0.2842	-0.1612	0.0360	9.90
02393	D	14336	0.5172	0.1800	0.0450	0.2560	0.5172	0.0019	0.2401	-0.2079	-0.2188	-0.2160	0.2401	1.9508	0.0301	9.90
02394	C	14336	0.6899	0.1467	0.0845	0.6899	0.0762	0.0027	0.3080	-0.2676	-0.2214	0.3080	-0.2109	0.6787	0.0323	9.90
02395	A	14273	0.6631	0.6631	0.0951	0.1467	0.0940	0.0012	0.3556	0.3556	-0.2449	-0.3525	-0.2278	0.8431	0.0317	9.90
02396	A	14273	0.7792	0.7792	0.0221	0.1126	0.0815	0.0046	0.2597	0.2597	-0.2228	-0.1577	-0.2113	-0.1120	0.0353	9.90
02397	D	14273	0.8679	0.0997	0.0177	0.0134	0.8679	0.0013	0.3546	-0.2775	-0.2174	-0.2260	0.3546	-1.1829	0.0428	6.90
02398	B	14273	0.7165	0.0640	0.7165	0.1292	0.0880	0.0022	0.3950	-0.2821	0.3950	-0.3098	-0.3054	0.3548	0.0333	9.90
02399	B	14273	0.4651	0.2033	0.4651	0.0975	0.2320	0.0022	0.0956	-0.1041	0.0956	-0.1858	-0.0125	2.2680	0.0304	9.90
02400	A	14273	0.9377	0.9377	0.0198	0.0287	0.0125	0.0013	0.3957	0.3957	-0.2658	-0.2770	-0.2041	-2.3673	0.0586	1.50
02401	C	14273	0.7173	0.0218	0.2433	0.7173	0.0161	0.0015	0.3838	-0.3067	-0.3284	0.3838	-0.2848	0.3178	0.0334	9.90
02402	C	14273	0.3402	0.1560	0.0277	0.3402	0.4730	0.0031	0.2758	-0.2846	-0.4489	0.2758	-0.2645	3.1693	0.0325	9.90
02403	C	14329	0.8105	0.0462	0.0754	0.8105	0.0664	0.0016	0.3027	-0.2291	-0.1058	0.3027	-0.3108	-0.3565	0.0374	9.90
02404	B	14329	0.2474	0.2434	0.2474	0.0589	0.4482	0.0022	0.1830	-0.0970	0.1830	-0.4570	-0.2160	4.1073	0.0362	9.90
02405	A	14329	0.6187	0.6187	0.0150	0.1208	0.2436	0.0018	0.1242	0.1242	-0.2518	-0.0866	-0.0861	1.2461	0.0312	9.90
02406	C	14329	0.7738	0.1091	0.0655	0.7738	0.0492	0.0024	0.2680	-0.1081	-0.2357	0.2680	-0.2738	0.0317	0.0353	9.90
02407	B	14329	0.8955	0.0385	0.8955	0.0301	0.0345	0.0014	0.4303	-0.2825	0.4303	-0.3068	-0.2697	-1.5205	0.0473	4.20
02408	D	14329	0.5297	0.0763	0.2706	0.1216	0.5297	0.0017	0.3753	-0.3983	-0.3167	-0.3417	0.3753	1.8772	0.0304	9.90
02409	D	14329	0.6856	0.1855	0.0780	0.0491	0.6856	0.0017	0.3709	-0.3074	-0.2994	-0.2620	0.3709	0.7559	0.0324	9.90
02410	A	14329	0.4144	0.4144	0.2619	0.2074	0.1142	0.0020	0.3547	0.3547	-0.1814	-0.5403	-0.3779	2.6914	0.0310	9.90
02411	A	14298	0.7222	0.7222	0.2358	0.0224	0.0189	0.0007	0.1316	0.1316	-0.0941	-0.1322	-0.1219	0.4472	0.0330	9.90
02412	B	14298	0.6445	0.1769	0.6445	0.0818	0.0955	0.0013	0.2122	-0.1095	0.2122	-0.2739	-0.1561	0.9958	0.0314	9.90
02413	A	14298	0.9621	0.9621	0.0103	0.0127	0.0140	0.0009	0.3100	0.3100	-0.1908	-0.1875	-0.1966	-3.0973	0.0751	1.00
02414	D	14298	0.9145	0.0202	0.0325	0.0315	0.9145	0.0013	0.3488	-0.2036	-0.2471	-0.2221	0.3488	-1.9255	0.0522	6.80
02415	B	14298	0.8696	0.0135	0.8696	0.0948	0.0212	0.0009	0.3039	-0.2123	0.3039	-0.2263	-0.1882	-1.1835	0.0432	9.90
02416	D	14298	0.8426	0.0287	0.0328	0.0947	0.8426	0.0012	0.4138	-0.2373	-0.3234	-0.3070	0.4138	-0.8179	0.0399	9.90
02417	C	14298	0.7323	0.0664	0.1390	0.7323	0.0603	0.0021	0.3033	-0.2037	-0.2784	0.3033	-0.1634	0.2827	0.0336	9.90
02418	B	14298	0.7030	0.1065	0.7030	0.0752	0.1123	0.0031	0.3978	-0.3117	0.3978	-0.2628	-0.3400	0.4812	0.0329	9.90
02419	B	14291	0.2085	0.5992	0.2085	0.0584	0.1322	0.0017	0.0979	-0.0782	0.0979	-0.1404	-0.2323	4.6804	0.0393	9.90
02420	D	14291	0.7656	0.0329	0.0395	0.1613	0.7656	0.0008	0.3779	-0.2391	-0.2826	-0.3072	0.3779	0.0498	0.0350	9.90
02421	B	14291	0.8414	0.0188	0.8414	0.0161	0.1228	0.0010	0.3597	-0.2467	0.3597	-0.2401	-0.2818	-0.8417	0.0403	9.90
02422	A	14291	0.7950	0.7950	0.0498	0.1201	0.0332	0.0020	0.3716	0.3716	-0.2995	-0.2708	-0.2375	-0.2586	0.0365	9.90
02423	D	14291	0.8712	0.0616	0.0572	0.0090	0.8712	0.0010	0.4149	-0.3754	-0.2418	-0.2011	0.4149	-1.2073	0.0434	4.00
02424	A	14291	0.6663	0.6663	0.1547	0.0733	0.1044	0.0013	0.4501	0.4501	-0.3748	-0.3822	-0.3410	0.8422	0.0321	9.90
02425	B	14291	0.2295	0.1660	0.2295	0.3328	0.2692	0.0025	0.0446	-0.0437	0.0446	-0.0732	-0.0269	4.3262	0.0374	9.90
02426	D	14291	0.7301	0.0435	0.1566	0.0676	0.7301	0.0022	0.3887	-0.3063	-0.3252	-0.2447	0.3887	0.3226	0.0338	9.90
02427	B	14304	0.7628	0.0229	0.7628	0.0219	0.1910	0.0014	0.3967	-0.2933	0.3967	-0.2103	-0.3474	-0.0052	0.0355	9.90
02428	B	14304	0.9115	0.0618	0.9115	0.0128	0.0131	0.0008	0.3334	-0.2559	0.3334	-0.2031	-0.1858	-1.7669	0.0503	8.00
02429	C	14304	0.5777	0.0582	0.3418	0.5777	0.0215	0.0007	0.2095	-0.2343	-0.1621	0.2095	-0.1975	1.5717	0.0306	9.90
02430	C	14304	0.8819	0.0652	0.0401	0.8819	0.0110	0.0019	0.4237	-0.2934	-0.3464	0.4237	-0.2015	-1.3416	0.0453	7.00
02431	B	14304	0.7043	0.1433	0.7043	0.1088	0.0423	0.0013	0.4027	-0.3309	0.4027	-0.2697	-0.3653	0.5025	0.0333	9.90

Item Detail			Proportions						Point Biserials					Rasch Statistics		
Item ID	Answer Key	N	P-Value	A	B	C	D	Other	Item Total Corr	A	B	C	D	Logit	SE	Outfit t
02432	A	14304	0.6252	0.6252	0.0844	0.0596	0.2299	0.0009	0.3561	0.3561	-0.3829	-0.3800	-0.2219	1.1644	0.0313	9.90
02433	C	14304	0.2768	0.2891	0.1536	0.2768	0.2763	0.0041	-0.0021	0.0287	-0.0716	-0.0021	0.0229	3.9580	0.0358	9.90
02434	D	14304	0.6465	0.0763	0.1311	0.1439	0.6465	0.0022	0.4485	-0.3828	-0.3561	-0.3732	0.4485	0.9438	0.0319	9.90
02435	A	14279	0.5428	0.5428	0.2304	0.0721	0.1526	0.0020	0.3076	0.3076	-0.2057	-0.2508	-0.3694	1.7827	0.0301	9.90
02436	B	14279	0.6173	0.1306	0.6173	0.1294	0.1214	0.0013	0.3094	-0.2450	0.3094	-0.2488	-0.2520	1.2922	0.0308	9.90
02437	D	14279	0.9493	0.0143	0.0161	0.0197	0.9493	0.0006	0.3409	-0.2138	-0.2435	-0.1872	0.3409	-2.5707	0.0646	8.90
02438	C	14279	0.7157	0.0460	0.1884	0.7157	0.0482	0.0017	0.3709	-0.3380	-0.2900	0.3709	-0.2425	0.4529	0.0333	9.90
02439	A	14279	0.4250	0.4250	0.3211	0.0992	0.1534	0.0013	0.1893	0.1893	-0.1532	-0.3151	-0.1349	2.6590	0.0307	9.90
02440	C	14279	0.8788	0.0358	0.0331	0.8788	0.0512	0.0011	0.3406	-0.2108	-0.1937	0.3406	-0.2602	-1.2719	0.0451	9.70
02441	C	14279	0.5311	0.0561	0.2451	0.5311	0.1658	0.0019	0.1457	-0.2499	-0.0928	0.1457	-0.1070	1.8611	0.0301	9.90
02442	A	14279	0.7138	0.7138	0.1125	0.0394	0.1319	0.0024	0.3087	0.3087	-0.2866	-0.3771	-0.1238	0.5569	0.0329	9.90
02443	A	14293	0.5129	0.5129	0.3427	0.0966	0.0469	0.0009	0.2231	0.2231	-0.1576	-0.2415	-0.3046	2.0119	0.0299	9.90
02444	D	14293	0.9043	0.0422	0.0297	0.0231	0.9043	0.0007	0.3814	-0.2622	-0.2255	-0.2635	0.3814	-1.5466	0.0479	8.00
02445	B	14293	0.7996	0.0696	0.7996	0.0812	0.0484	0.0013	0.2822	-0.1737	0.2822	-0.2134	-0.2015	-0.2886	0.0367	9.90
02446	C	14293	0.5129	0.1771	0.1021	0.5129	0.2056	0.0023	0.2086	-0.1163	-0.2653	0.2086	-0.1946	2.0155	0.0299	9.90
02447	A	14293	0.7466	0.7466	0.1321	0.0763	0.0437	0.0014	0.3406	0.3406	-0.2722	-0.2342	-0.2447	0.1849	0.0341	9.90
02448	B	14293	0.7525	0.0352	0.7525	0.0638	0.1474	0.0010	0.3246	-0.2495	0.3246	-0.3649	-0.1708	0.1161	0.0345	9.90
02449	D	14293	0.9055	0.0092	0.0373	0.0470	0.9055	0.0010	0.3418	-0.2229	-0.2340	-0.2298	0.3418	-1.7158	0.0501	7.50
02450	B	14293	0.6609	0.1478	0.6609	0.0623	0.1272	0.0018	0.3632	-0.2645	0.3632	-0.2123	-0.3641	0.8879	0.0315	9.90
02451	C	14262	0.8220	0.0211	0.1289	0.8220	0.0266	0.0013	0.4009	-0.2716	-0.3279	0.4009	-0.2410	-0.6083	0.0386	9.90
02452	B	14262	0.7492	0.0768	0.7492	0.0810	0.0916	0.0014	0.4128	-0.3295	0.4128	-0.3393	-0.2599	0.0914	0.0346	9.90
02453	D	14262	0.8453	0.0839	0.0475	0.0224	0.8453	0.0009	0.4426	-0.3316	-0.3323	-0.2573	0.4426	-0.7964	0.0400	9.90
02454	C	14262	0.8397	0.0691	0.0741	0.8397	0.0158	0.0013	0.3444	-0.1900	-0.2983	0.3444	-0.2334	-0.8109	0.0401	9.90
02455	A	14262	0.7033	0.7033	0.0797	0.1835	0.0322	0.0013	0.3377	0.3377	-0.3473	-0.2301	-0.2233	0.5482	0.0328	9.90
02456	B	14262	0.5656	0.2386	0.5656	0.1043	0.0886	0.0028	0.3200	-0.2826	0.3200	-0.3224	-0.2052	1.5500	0.0305	9.90
02457	C	14262	0.7774	0.1299	0.0686	0.7774	0.0225	0.0016	0.3432	-0.2338	-0.3101	0.3432	-0.2040	-0.1653	0.0359	9.90
02458	D	14262	0.8125	0.0275	0.0393	0.1183	0.8125	0.0024	0.3526	-0.2390	-0.2794	-0.2496	0.3526	-0.4804	0.0378	9.90
02459	C	14295	0.7119	0.0163	0.1177	0.7119	0.1531	0.0009	0.4632	-0.3052	-0.4428	0.4632	-0.3370	0.4332	0.0331	9.90
02460	B	14295	0.7307	0.1970	0.7307	0.0277	0.0439	0.0006	0.2718	-0.1375	0.2718	-0.2725	-0.3495	0.2759	0.0337	9.90
02461	C	14295	0.4318	0.1671	0.2787	0.4318	0.1210	0.0015	0.2710	-0.2851	-0.2335	0.2710	-0.2937	2.5559	0.0306	9.90
02462	D	14295	0.8651	0.0627	0.0400	0.0297	0.8651	0.0025	0.3761	-0.1805	-0.3272	-0.2877	0.3761	-1.0980	0.0426	9.90
02463	A	14295	0.8260	0.8260	0.0537	0.1014	0.0178	0.0012	0.4066	0.4066	-0.3046	-0.3325	-0.1771	-0.5971	0.0385	9.90
02464	B	14295	0.5668	0.2072	0.5668	0.1406	0.0835	0.0018	0.2229	-0.2325	0.2229	-0.0624	-0.2836	1.5776	0.0303	9.90
02465	A	14295	0.7817	0.7817	0.1593	0.0312	0.0267	0.0012	0.3681	0.3681	-0.2971	-0.2668	-0.2502	-0.1381	0.0357	9.90
02466	C	14295	0.7838	0.0716	0.0931	0.7838	0.0493	0.0021	0.4382	-0.3751	-0.2903	0.4382	-0.3084	-0.2493	0.0363	9.90

Appendix G:

2006 Uncommon Grade 11 Multiple Choice Statistics for Writing

Item Detail			Proportions						Point Biserials					Rasch Statistics		
Item ID	Answer Key	N	P-Value	A	B	C	D	Other	Item Total Corr	A	B	C	D	Logit	SE	Outfit t
02467	A	13312	0.7516	0.7516	0.0430	0.0565	0.1482	0.0007	0.2575	0.2575	-0.2363	-0.2192	-0.1579	0.5650	0.0368	9.90
02468	B	13312	0.6498	0.2747	0.6498	0.0587	0.0163	0.0005	0.2598	-0.2509	0.2598	-0.1549	-0.1276	1.3948	0.0337	9.90
02469	B	13312	0.1873	0.0739	0.1873	0.4161	0.3210	0.0017	0.1584	-0.3038	0.1584	-0.1089	-0.2723	5.2451	0.0436	9.90
02470	D	13312	0.8999	0.0169	0.0391	0.0438	0.8999	0.0003	0.2805	-0.1395	-0.1523	-0.2398	0.2805	-1.1689	0.0502	9.90
02471	C	13312	0.7505	0.0889	0.0704	0.7505	0.0887	0.0016	0.3184	-0.3331	-0.2170	0.3184	-0.1492	0.5867	0.0367	9.90
02472	D	13312	0.5891	0.0132	0.1999	0.1961	0.5891	0.0017	0.3498	-0.2922	-0.2710	-0.3585	0.3498	1.8597	0.0327	9.90
02473	C	13312	0.5297	0.2992	0.0355	0.5297	0.1342	0.0013	0.3140	-0.3230	-0.3146	0.3140	-0.1931	2.2573	0.0322	9.90
02474	A	13312	0.3187	0.3187	0.1699	0.1827	0.3198	0.0089	0.1824	0.1824	-0.3273	0.0288	-0.2433	3.8875	0.0353	9.90
02475	D	13182	0.8734	0.0460	0.0669	0.0132	0.8734	0.0005	0.4012	-0.2617	-0.3267	-0.2118	0.4012	-0.8003	0.0466	9.90
02476	C	13182	0.4878	0.0333	0.2733	0.4878	0.2047	0.0010	0.0652	-0.1821	0.0549	0.0652	-0.1551	2.6741	0.0323	9.90
02477	D	13182	0.6336	0.0194	0.0479	0.2980	0.6336	0.0011	0.4230	-0.3552	-0.3849	-0.3675	0.4230	1.5113	0.0334	9.90
02478	C	13182	0.5449	0.0702	0.1478	0.5449	0.2355	0.0016	0.2578	-0.2623	-0.2717	0.2578	-0.1835	2.1510	0.0323	9.90
02479	A	13182	0.7392	0.7392	0.0456	0.1179	0.0962	0.0011	0.4661	0.4661	-0.3475	-0.3486	-0.3986	0.5465	0.0369	9.90
02480	B	13182	0.9015	0.0322	0.9015	0.0434	0.0222	0.0008	0.4546	-0.3160	0.4546	-0.3411	-0.2412	-1.3719	0.0534	1.80
02481	C	13182	0.5329	0.0385	0.2274	0.5329	0.1997	0.0015	0.2739	-0.2655	-0.2264	0.2739	-0.2568	2.1900	0.0323	9.90
02482	B	13182	0.8066	0.0863	0.8066	0.0823	0.0229	0.0018	0.3570	-0.2310	0.3570	-0.2916	-0.2418	-0.0254	0.0402	9.90
02483	B	13201	0.9470	0.0137	0.9470	0.0230	0.0159	0.0003	0.3516	-0.2217	0.3516	-0.2498	-0.1847	-2.2959	0.0695	3.20
02484	D	13201	0.7509	0.1386	0.0451	0.0648	0.7509	0.0005	0.4479	-0.2935	-0.2935	-0.4781	0.4479	0.4614	0.0373	9.90
02485	B	13201	0.8931	0.0298	0.8931	0.0526	0.0239	0.0007	0.3373	-0.2572	0.3373	-0.2141	-0.1987	-1.1310	0.0500	7.50
02486	A	13201	0.8701	0.8701	0.0248	0.0522	0.0522	0.0007	0.4104	0.4104	-0.2423	-0.3422	-0.2461	-0.7678	0.0460	9.90
02487	B	13201	0.8959	0.0654	0.8959	0.0332	0.0047	0.0008	0.3278	-0.2345	0.3278	-0.2477	-0.1779	-1.2002	0.0509	9.90
02488	A	13201	0.7059	0.7059	0.0321	0.2370	0.0239	0.0010	0.2577	0.2577	-0.3051	-0.1683	-0.2692	0.9291	0.0353	9.90
02489	D	13201	0.8864	0.0398	0.0372	0.0345	0.8864	0.0020	0.4117	-0.2293	-0.2954	-0.3035	0.4117	-1.0181	0.0487	5.10
02490	C	13201	0.7937	0.0234	0.1579	0.7937	0.0236	0.0014	0.2491	-0.1486	-0.1825	0.2491	-0.2350	0.2095	0.0386	9.90
02491	B	13177	0.8804	0.0582	0.8804	0.0371	0.0240	0.0003	0.3370	-0.2076	0.3370	-0.3226	-0.1389	-0.9215	0.0472	9.90
02492	A	13177	0.9366	0.9366	0.0093	0.0168	0.0370	0.0003	0.4018	0.4018	-0.2306	-0.2268	-0.3069	-2.0753	0.0636	2.60
02493	C	13177	0.6640	0.0726	0.0924	0.6640	0.1693	0.0017	0.3587	-0.3342	-0.3073	0.3587	-0.2518	1.2012	0.0345	9.90
02494	B	13177	0.7537	0.0500	0.7537	0.1137	0.0822	0.0004	0.3880	-0.2762	0.3880	-0.3446	-0.2365	0.3586	0.0377	9.90
02495	C	13177	0.8986	0.0609	0.0247	0.8986	0.0146	0.0011	0.3908	-0.2731	-0.3033	0.3908	-0.1975	-1.2351	0.0508	8.00
02496	A	13177	0.9663	0.9663	0.0146	0.0137	0.0046	0.0008	0.2725	0.2725	-0.1962	-0.1726	-0.1150	-2.9570	0.0837	2.80
02497	C	13177	0.4616	0.1143	0.1524	0.4616	0.2680	0.0037	0.2570	-0.1022	-0.2418	0.2570	-0.3033	2.7681	0.0327	9.90
02498	B	13177	0.8151	0.0693	0.8151	0.0691	0.0447	0.0018	0.3523	-0.2599	0.3523	-0.2463	-0.2316	-0.2047	0.0411	9.90
02499	C	13244	0.8574	0.0379	0.0570	0.8574	0.0471	0.0006	0.3279	-0.1123	-0.2530	0.3279	-0.2812	-0.6735	0.0450	9.90
02500	A	13244	0.7225	0.7225	0.1273	0.1059	0.0421	0.0022	0.3687	0.3687	-0.3808	-0.2100	-0.2150	0.7186	0.0359	9.90
02501	B	13244	0.5926	0.3435	0.5926	0.0370	0.0264	0.0005	0.3779	-0.3485	0.3779	-0.2939	-0.2844	1.7550	0.0329	9.90
02502	D	13244	0.9500	0.0102	0.0172	0.0219	0.9500	0.0007	0.3483	-0.2206	-0.2119	-0.2222	0.3483	-2.3337	0.0708	1.60
02503	B	13244	0.7733	0.0835	0.7733	0.0769	0.0650	0.0012	0.4207	-0.3016	0.4207	-0.3205	-0.3089	0.3449	0.0376	9.90
02504	C	13244	0.8641	0.0371	0.0823	0.8641	0.0161	0.0005	0.3738	-0.2312	-0.2988	0.3738	-0.2184	-0.7213	0.0455	6.10
02505	C	13244	0.5371	0.1560	0.2103	0.5371	0.0940	0.0026	0.2162	-0.2014	-0.1613	0.2162	-0.2104	2.3212	0.0323	9.90
02506	A	13244	0.7983	0.7983	0.0675	0.1064	0.0261	0.0017	0.3133	0.3133	-0.2294	-0.2071	-0.2671	0.1239	0.0389	9.90
02507	B	13214	0.7170	0.0479	0.7170	0.0639	0.1707	0.0005	0.4052	-0.2534	0.4052	-0.3021	-0.3591	0.8091	0.0355	9.90
02508	A	13214	0.8163	0.8163	0.0598	0.0911	0.0325	0.0003	0.4182	0.4182	-0.2969	-0.3420	-0.2342	-0.1363	0.0405	9.90
02509	A	13214	0.6114	0.6114	0.0736	0.2240	0.0900	0.0010	0.2471	0.2471	-0.3992	-0.0898	-0.2504	1.7196	0.0330	9.90
02510	C	13214	0.8282	0.0626	0.0245	0.8282	0.0835	0.0011	0.3478	-0.2484	-0.2353	0.3478	-0.2516	-0.2258	0.0412	9.90
02511	C	13214	0.9274	0.0299	0.0304	0.9274	0.0114	0.0009	0.3844	-0.2466	-0.2833	0.3844	-0.1976	-1.8152	0.0600	2.70

Item Detail			Proportions						Point Biserials					Rasch Statistics		
Item ID	Answer Key	N	P-Value	A	B	C	D	Other	Item Total Corr	A	B	C	D	Logit	SE	Outfit t
02512	A	13214	0.9161	0.9161	0.0362	0.0313	0.0156	0.0008	0.3514	0.3514	-0.2255	-0.2296	-0.2310	-1.5750	0.0561	6.90
02513	B	13214	0.7190	0.0817	0.7190	0.1013	0.0965	0.0015	0.4444	-0.3665	0.4444	-0.3292	-0.3372	0.7476	0.0357	9.90
02514	D	13214	0.7330	0.1738	0.0381	0.0541	0.7330	0.0010	0.1367	0.0125	-0.2136	-0.2392	0.1367	0.7554	0.0357	9.90
02515	C	13195	0.7036	0.0122	0.2688	0.7036	0.0146	0.0008	0.1672	-0.2265	-0.1147	0.1672	-0.2037	1.0316	0.0348	9.90
02516	A	13195	0.5636	0.5636	0.0488	0.1108	0.2755	0.0013	0.2781	0.2781	-0.3465	-0.2609	-0.2050	2.0408	0.0327	9.90
02517	C	13195	0.9042	0.0569	0.0180	0.9042	0.0202	0.0007	0.2623	-0.1818	-0.1550	0.2623	-0.1712	-1.2708	0.0513	9.90
02518	B	13195	0.7792	0.1905	0.7792	0.0137	0.0161	0.0005	0.3997	-0.3723	0.3997	-0.1866	-0.2109	0.1777	0.0385	9.90
02519	B	13195	0.8961	0.0584	0.8961	0.0293	0.0150	0.0012	0.2976	-0.2334	0.2976	-0.1753	-0.1479	-1.0855	0.0490	9.90
02520	C	13195	0.6809	0.1141	0.0863	0.6809	0.1171	0.0017	0.4166	-0.3281	-0.3588	0.4166	-0.3278	1.0863	0.0346	9.90
02521	A	13195	0.8139	0.8139	0.0429	0.0642	0.0775	0.0016	0.4608	0.4608	-0.3449	-0.3376	-0.3277	-0.0801	0.0400	9.90
02522	D	13195	0.7388	0.0490	0.1629	0.0481	0.7388	0.0011	0.4174	-0.2444	-0.3779	-0.2916	0.4174	0.6018	0.0364	9.90
02523	A	13239	0.5278	0.5278	0.1312	0.1713	0.1685	0.0011	0.3077	0.3077	-0.1994	-0.3017	-0.3213	2.2939	0.0324	9.90
02524	D	13239	0.6281	0.0564	0.2968	0.0183	0.6281	0.0004	0.3518	-0.4387	-0.2709	-0.2663	0.3518	1.5486	0.0333	9.90
02525	C	13239	0.6455	0.0292	0.0357	0.6455	0.2888	0.0008	0.4025	-0.2818	-0.3673	0.4025	-0.3555	1.3701	0.0337	9.90
02526	B	13239	0.7614	0.1548	0.7614	0.0326	0.0505	0.0008	0.2942	-0.1995	0.2942	-0.2301	-0.2513	0.4510	0.0371	9.90
02527	A	13239	0.8616	0.8616	0.0632	0.0607	0.0131	0.0014	0.4317	0.4317	-0.2662	-0.3772	-0.2377	-0.5944	0.0444	6.60
02528	A	13239	0.7461	0.7461	0.0216	0.0361	0.1947	0.0014	0.3729	0.3729	-0.2350	-0.3899	-0.2817	0.6008	0.0364	9.90
02529	C	13239	0.6999	0.0303	0.0242	0.6999	0.2441	0.0016	0.4110	-0.3770	-0.3543	0.4110	-0.3301	0.9220	0.0351	9.90
02530	B	13239	0.6250	0.3426	0.6250	0.0182	0.0111	0.0031	0.4139	-0.4192	0.4139	-0.1430	-0.1106	1.5296	0.0333	9.90
02531	D	13256	0.7693	0.0216	0.1209	0.0874	0.7693	0.0008	0.2596	-0.1250	-0.1814	-0.2311	0.2596	0.3991	0.0370	9.90
02532	B	13256	0.8724	0.0268	0.8724	0.0346	0.0656	0.0007	0.4270	-0.2028	0.4270	-0.2645	-0.3699	-0.7902	0.0459	9.90
02533	C	13256	0.4516	0.2464	0.1981	0.4516	0.1010	0.0029	0.2523	-0.2424	-0.2048	0.2523	-0.2937	2.7666	0.0324	9.90
02534	C	13256	0.7904	0.0714	0.0512	0.7904	0.0861	0.0009	0.2307	-0.2559	-0.2259	0.2307	-0.0266	0.2070	0.0381	9.90
02535	B	13256	0.7983	0.1377	0.7983	0.0357	0.0275	0.0008	0.3487	-0.2428	0.3487	-0.3089	-0.2402	0.0682	0.0389	9.90
02536	B	13256	0.8235	0.0124	0.8235	0.0499	0.1128	0.0015	0.3269	-0.1953	0.3269	-0.2933	-0.2170	-0.1652	0.0405	9.90
02537	C	13256	0.8868	0.0245	0.0717	0.8868	0.0154	0.0016	0.3519	-0.2312	-0.2616	0.3519	-0.2004	-1.0377	0.0486	9.60
02538	A	13256	0.7120	0.7120	0.1288	0.1357	0.0210	0.0026	0.2630	0.2630	-0.2116	-0.1746	-0.2791	0.9089	0.0349	9.90
02539	B	13190	0.9469	0.0205	0.9469	0.0219	0.0103	0.0003	0.3418	-0.2220	0.3418	-0.2362	-0.1779	-2.1213	0.0657	1.20
02540	A	13190	0.9562	0.9562	0.0090	0.0061	0.0283	0.0005	0.3074	0.3074	-0.1758	-0.1672	-0.2247	-2.5801	0.0758	4.20
02541	C	13190	0.6466	0.1691	0.0737	0.6466	0.1096	0.0011	0.3574	-0.2954	-0.3792	0.3574	-0.2183	1.3934	0.0338	9.90
02542	C	13190	0.7882	0.1195	0.0529	0.7882	0.0386	0.0008	0.3815	-0.2848	-0.2289	0.3815	-0.3379	0.1766	0.0388	9.90
02543	A	13190	0.1708	0.1708	0.0956	0.6541	0.0782	0.0013	0.0993	0.0993	-0.2809	-0.0771	-0.2165	5.6645	0.0466	9.90
02544	B	13190	0.6572	0.0409	0.6572	0.2782	0.0222	0.0016	0.1274	-0.2052	0.1274	-0.0551	-0.1865	1.4315	0.0337	9.90
02545	D	13190	0.7888	0.0620	0.0634	0.0831	0.7888	0.0027	0.4946	-0.3422	-0.3678	-0.3967	0.4946	0.2083	0.0386	7.30
02546	A	13190	0.6136	0.6136	0.0402	0.1470	0.1970	0.0023	0.0091	0.0091	-0.3040	0.0962	0.0245	1.7955	0.0330	9.90

Appendix H:

2006 Common Grade 5 Multiple Choice Statistics for Writing

Item Detail			Proportions						Point Biserials					Rasch Statistics		
Item ID	Answer Key	N	P-Value	A	B	C	D	Other	Item Total Corr	A	B	C	D	Logit	SE	Outfit t
02265	B	131463	0.7608	0.1259	0.7608	0.0505	0.0614	0.0014	0.3437	-0.2075	0.3437	-0.3131	-0.2743	0.1320	0.0113	9.90
02266	B	131463	0.7776	0.1482	0.7776	0.0326	0.0403	0.0012	0.4310	-0.3549	0.4310	-0.3080	-0.2777	-0.0175	0.0116	9.90
02267	D	131463	0.7352	0.0704	0.0385	0.1542	0.7352	0.0017	0.2843	-0.1656	-0.3014	-0.2067	0.2843	0.3600	0.0110	9.90
02268	C	131463	0.7797	0.1369	0.0382	0.7797	0.0434	0.0019	0.2407	-0.1100	-0.1931	0.2407	-0.2788	0.0608	0.0115	9.90
02269	A	131463	0.8184	0.8184	0.0407	0.0940	0.0453	0.0016	0.3136	0.3136	-0.2575	-0.1802	-0.2337	-0.2526	0.0121	9.90
02270	A	131463	0.8892	0.8892	0.0507	0.0296	0.0291	0.0015	0.4084	0.4084	-0.2566	-0.3040	-0.2613	-0.9674	0.0140	9.90
02271	C	131463	0.9226	0.0334	0.0276	0.9226	0.0147	0.0017	0.3672	-0.2707	-0.2251	0.3672	-0.1948	-1.3887	0.0157	9.90
02272	D	131463	0.7994	0.0296	0.0974	0.0714	0.7994	0.0022	0.4166	-0.2878	-0.3030	-0.3104	0.4166	-0.1279	0.0118	9.90
02273	B	131463	0.5137	0.4021	0.5137	0.0257	0.0570	0.0016	0.2143	-0.1624	0.2143	-0.3524	-0.2091	1.6483	0.0100	9.90
02274	B	131463	0.8297	0.0307	0.8297	0.0704	0.0673	0.0019	0.4139	-0.2236	0.4139	-0.3370	-0.2820	-0.3811	0.0124	9.90
02275	C	131463	0.5807	0.2881	0.0440	0.5807	0.0847	0.0025	0.3616	-0.3523	-0.3029	0.3616	-0.2216	1.2510	0.0102	9.90
02276	D	131463	0.8297	0.0868	0.0357	0.0448	0.8297	0.0030	0.3154	-0.1914	-0.2439	-0.2326	0.3154	-0.3168	0.0122	9.90

Appendix I:

2006 Common Grade 8 Multiple Choice Statistics for Writing

Item Detail			Proportions						Point Biserials					Rasch Statistics		
Item ID	Answer Key	N	P-Value	A	B	C	D	Other	Item Total Corr	A	B	C	D	Logit	SE	Outfit t
02277	A	142960	0.8562	0.8562	0.0765	0.0425	0.0236	0.0012	0.2772	0.2772	-0.2119	-0.1739	-0.1567	-0.6759	0.0124	9.90
02278	B	142960	0.7046	0.0162	0.7046	0.2341	0.0437	0.0014	0.3880	-0.2298	0.3880	-0.3460	-0.2787	0.3722	0.0106	9.90
02279	B	142960	0.4177	0.3889	0.4177	0.1246	0.0676	0.0013	0.3022	-0.2346	0.3022	-0.4602	-0.2996	2.0356	0.0096	9.90
02280	C	142960	0.8809	0.0217	0.0090	0.8809	0.0869	0.0014	0.3054	-0.1843	-0.1616	0.3054	-0.2436	-0.9472	0.0131	9.90
02281	C	142960	0.8691	0.0689	0.0367	0.8691	0.0244	0.0009	0.4014	-0.3060	-0.2722	0.4014	-0.2184	-0.8730	0.0129	9.90
02282	C	142960	0.6142	0.0845	0.1057	0.6142	0.1946	0.0010	0.2866	-0.0946	-0.1276	0.2866	-0.3775	0.9335	0.0100	9.90
02283	D	142960	0.8461	0.0134	0.0554	0.0839	0.8461	0.0012	0.3199	-0.2194	-0.2029	-0.2431	0.3199	-0.6110	0.0123	9.90
02284	A	142960	0.8133	0.8133	0.1405	0.0175	0.0262	0.0025	0.2650	0.2650	-0.1921	-0.2080	-0.1901	-0.3257	0.0117	9.90
02285	D	142960	0.6483	0.0782	0.2097	0.0622	0.6483	0.0015	0.3792	-0.3737	-0.2943	-0.2868	0.3792	0.7224	0.0102	9.90
02286	A	142960	0.7408	0.7408	0.0861	0.0731	0.0972	0.0029	0.3966	0.3966	-0.2965	-0.3038	-0.2910	0.1583	0.0109	9.90
02287	C	142960	0.7846	0.0307	0.1377	0.7846	0.0452	0.0018	0.3436	-0.2430	-0.2379	0.3436	-0.2915	-0.1050	0.0113	9.90
02288	A	142960	0.8484	0.8484	0.0330	0.0252	0.0915	0.0020	0.4083	0.4083	-0.2976	-0.2678	-0.2927	-0.6841	0.0124	9.90

Appendix J:

2006 Common Grade 11 Multiple Choice Statistics for Writing

Item Detail			Proportions						Point Biserials					Rasch Statistics		
Item ID	Answer Key	N	P-Value	A	B	C	D	Other	Item Total Corr	A	B	C	D	Logit	SE	Outfit t
02289	A	132210	0.9097	0.9097	0.0270	0.0472	0.0152	0.0010	0.3821	0.3821	-0.2659	-0.3064	-0.1303	-1.1036	0.0157	9.90
02290	B	132210	0.9024	0.0168	0.9024	0.0606	0.0198	0.0005	0.2825	-0.1998	0.2825	-0.1716	-0.2083	-0.9235	0.0150	9.90
02291	D	132210	0.7546	0.0975	0.1052	0.0416	0.7546	0.0011	0.4065	-0.3343	-0.2471	-0.3780	0.4065	0.4183	0.0118	9.90
02292	C	132210	0.6412	0.0454	0.0993	0.6412	0.2131	0.0011	0.3886	-0.4407	-0.4069	0.3886	-0.2431	1.1877	0.0108	9.90
02293	A	132210	0.6864	0.6864	0.0242	0.1412	0.1473	0.0009	0.3490	0.3490	-0.2702	-0.3398	-0.2264	0.8996	0.0112	9.90
02294	C	132210	0.4918	0.1028	0.3503	0.4918	0.0535	0.0015	0.2798	-0.2922	-0.2314	0.2798	-0.3326	2.1149	0.0103	9.90
02295	B	132210	0.8330	0.0639	0.8330	0.0398	0.0613	0.0019	0.3691	-0.2333	0.3691	-0.2734	-0.2662	-0.1778	0.0129	9.90
02296	B	132210	0.9292	0.0359	0.9292	0.0178	0.0157	0.0015	0.3017	-0.2521	0.3017	-0.1721	-0.1195	-1.3568	0.0167	9.90
02297	C	132210	0.8809	0.0649	0.0365	0.8809	0.0167	0.0010	0.4140	-0.2934	-0.3551	0.4140	-0.1553	-0.6899	0.0143	9.90
02298	A	132210	0.6844	0.6844	0.1749	0.0864	0.0526	0.0017	0.4535	0.4535	-0.4292	-0.4085	-0.1666	0.8728	0.0112	9.90
02299	C	132210	0.8979	0.0242	0.0530	0.8979	0.0237	0.0012	0.4218	-0.2672	-0.3043	0.4218	-0.2620	-0.9201	0.0150	9.90
02300	D	132210	0.8460	0.0714	0.0141	0.0669	0.8460	0.0016	0.3707	-0.2906	-0.2332	-0.2448	0.3707	-0.3216	0.0133	9.90

Appendix K:

2006 Common Grade 5 Constructed Response Statistics for Writing

Item Detail					Proportions					Point Biserials					Rasch Statistics		
Item ID	Max Score Points	Item Weight	Prompt Domain	N	P-Value	1	2	3	4	Item Total Corr	1	2	3	4	Logit	SE	Outfit t
02301	4	10	Composition	131463	0.6859	0.0185	0.2993	0.6024	0.0798	0.5726	-0.2723	-0.5082	-0.3062	0.3062	1.2045	0.0088	-9.90
02301	4	1	Revising & Editing	131463	0.6853	0.0241	0.3011	0.5843	0.0905	0.6658	-0.3340	-0.5659	-0.4033	0.4033	1.3734	0.0085	9.90
02302	4	10	Composition	131463	0.6612	0.0239	0.3647	0.5540	0.0574	0.5686	-0.3000	-0.5048	-0.2749	0.2749	1.6629	0.0087	-9.90
02302	4	1	Revising & Editing	131463	0.6766	0.0269	0.3094	0.5938	0.0699	0.6602	-0.3514	-0.5674	-0.3652	0.3652	1.6212	0.0086	9.90

NOTE: Overall P-value is an indicator of item difficulty, with higher values indicating easier items. Category proportion values are the percentage of students attaining each score category.

Appendix L:

2006 Common Grade 8 Constructed Response Statistics for Writing

Item Detail					Proportions					Point Biserials					Rasch Statistics		
Item ID	Max Score Points	Item Weight	Prompt Domain	N	P-Value	1	2	3	4	Item Total Corr	1	2	3	4	Logit	SE	Outfit t
02303	4	10	Composition	142960	0.7156	0.0224	0.2236	0.6232	0.1308	0.6573	-0.3306	-0.5875	-0.3755	0.3755	0.4623	0.0085	-9.90
02303	4	1	Revising & Editing	142960	0.7120	0.0217	0.2406	0.6059	0.1319	0.7229	-0.3697	-0.6224	-0.4425	0.4425	0.5563	0.0084	9.90
02304	4	10	Composition	142960	0.7078	0.0252	0.2417	0.6098	0.1233	0.6497	-0.3405	-0.5788	-0.3658	0.3658	0.6628	0.0084	-9.90
02304	4	1	Revising & Editing	142960	0.7053	0.0244	0.2510	0.6038	0.1208	0.7240	-0.3849	-0.6252	-0.4300	0.4300	0.7269	0.0084	9.90

NOTE: Overall P-value is an indicator of item difficulty, with higher values indicating easier items. Category proportion values are the percentage of students attaining each score category.

Appendix M:

2006 Common Grade 11 Constructed Response Statistics for Writing

Item Detail					Proportions					Point Biserials					Rasch Statistics		
Item ID	Max Score Points	Item Weight	Prompt Domain	N	P-Value	1	2	3	4	Item Total Corr	1	2	3	4	Logit	SE	Outfit t
02305	4	10	Composition	132210	0.7096	0.0253	0.2233	0.6388	0.1125	0.6274	-0.3536	-0.5475	-0.3483	0.3483	1.3795	0.0090	-9.90
02305	4	1	Revising & Editing	132210	0.7099	0.0222	0.2327	0.6284	0.1168	0.7356	-0.3990	-0.6252	-0.4397	0.4397	1.3541	0.0089	9.90
02306	4	10	Composition	132210	0.7083	0.0200	0.2161	0.6746	0.0893	0.6137	-0.3216	-0.5399	-0.3248	0.3248	1.3041	0.0095	-9.90
02306	4	1	Revising & Editing	132210	0.7054	0.0183	0.2385	0.6465	0.0967	0.7132	-0.3713	-0.6132	-0.4062	0.4062	1.3680	0.0092	9.90

NOTE: Overall P-value is an indicator of item difficulty, with higher values indicating easier items. Category proportion values are the percentage of students attaining each score category.

Appendix N:

2006 Raw to Scale Score Tables

WRITING GRADE 5

Raw Score	Measure	Scale Score	Logit SE	Scale Score SE	Freq	Freq %	Cum Freq	Cum Freq %	Percentile
22	-7.8104	700	1.838	184	6	0	6	0	1
23	-6.5745	700	1.0228	102	28	0	34	0	1
24	-5.8343	700	0.7403	74	94	0.1	128	0.1	1
25	-5.3797	700	0.6193	62	153	0.1	281	0.2	1
26	-5.0408	700	0.5499	55	192	0.1	473	0.4	1
27	-4.7641	700	0.5048	50	164	0.1	637	0.5	1
28	-4.5256	700	0.4735	47	141	0.1	778	0.6	1
29	-4.3125	700	0.4508	45	112	0.1	890	0.7	1
30	-4.1171	700	0.4342	43	54	0	944	0.7	1
31	-3.9341	700	0.422	42	32	0	976	0.7	1
32	-3.7599	700	0.4132	41	23	0	999	0.8	1
33	-3.5918	712	0.4071	41	30	0	1029	0.8	1
34	-3.4277	729	0.4033	40	88	0.1	1117	0.8	1
35	-3.266	745	0.4014	40	199	0.2	1316	1	1
36	-3.105	761	0.4011	40	286	0.2	1602	1.2	1
37	-2.9438	777	0.4022	40	428	0.3	2030	1.5	1
38	-2.7813	793	0.4044	40	439	0.3	2469	1.9	2
39	-2.6166	810	0.4073	41	420	0.3	2889	2.2	2
40	-2.4493	827	0.4107	41	368	0.3	3257	2.5	2
41	-2.2792	844	0.4141	41	301	0.2	3558	2.7	3
42	-2.1064	861	0.4171	42	240	0.2	3798	2.9	3
43	-1.9316	878	0.419	42	207	0.2	4005	3	3
44	-1.7557	896	0.4195	42	183	0.1	4188	3.2	3
45	-1.58	913	0.4184	42	275	0.2	4463	3.4	3
46	-1.406	931	0.4156	42	490	0.4	4953	3.8	4
47	-1.235	948	0.4113	41	906	0.7	5859	4.5	4
48	-1.0679	965	0.4059	41	1412	1.1	7271	5.5	5
49	-0.9057	981	0.3997	40	1947	1.5	9218	7	6
50	-0.7486	997	0.3931	39	2341	1.8	11559	8.8	8
51	-0.5966	1012	0.3866	39	2732	2.1	14291	10.9	10
52	-0.4495	1026	0.3803	38	2976	2.3	17267	13.1	12
53	-0.3072	1041	0.3745	37	2950	2.2	20217	15.4	14
54	-0.1689	1055	0.3692	37	2815	2.1	23032	17.5	16
55	-0.0344	1068	0.3646	36	2501	1.9	25533	19.4	18
56	0.0971	1081	0.3606	36	2010	1.5	27543	21	20
57	0.226	1094	0.3574	36	1519	1.2	29062	22.1	22
58	0.3528	1107	0.355	36	1028	0.8	30090	22.9	22
59	0.4782	1119	0.3533	35	918	0.7	31008	23.6	23
60	0.6027	1132	0.3525	35	1534	1.2	32542	24.8	24
61	0.7269	1144	0.3524	35	2119	1.6	34661	26.4	26
62	0.8513	1157	0.3533	35	2980	2.3	37641	28.6	27
63	0.9767	1169	0.3551	36	3705	2.8	41346	31.5	30
64	1.1037	1182	0.3579	36	4544	3.5	45890	34.9	33
65	1.2332	1195	0.3618	36	5063	3.9	50953	38.8	37
66	1.3659	1208	0.367	37	5122	3.9	56075	42.7	41
67	1.5029	1222	0.3737	37	4438	3.4	60513	46	44
68	1.6456	1236	0.3821	38	2554	1.9	63067	48	47
69	1.7955	1251	0.3924	39	482	0.4	63549	48.3	48
70	1.9544	1267	0.4053	41	681	0.5	64230	48.9	49
71	2.1249	1284	0.4211	42	1157	0.9	65387	49.7	49
72	2.3103	1302	0.4407	44	2105	1.6	67492	51.3	51
73	2.515	1323	0.465	47	3183	2.4	70675	53.8	53
74	2.745	1346	0.4951	50	5129	3.9	75804	57.7	56
75	3.0081	1372	0.5316	53	7553	5.7	83357	63.4	61
76	3.3129	1403	0.573	57	10147	7.7	93504	71.1	67
77	3.6648	1438	0.6114	61	11741	8.9	105245	80.1	76
78	4.0531	1477	0.6301	63	9445	7.2	114690	87.2	84
79	4.4454	1516	0.6175	62	2226	1.7	116916	88.9	88
80	4.8074	1552	0.5838	58	291	0.2	117207	89.2	89
81	5.1265	1584	0.5462	55	38	0	117245	89.2	89
82	5.4067	1612	0.5132	51	68	0.1	117313	89.2	89
83	5.6561	1637	0.4866	49	156	0.1	117469	89.4	89
84	5.8827	1660	0.4662	47	318	0.2	117787	89.6	89
85	6.0925	1681	0.4508	45	674	0.5	118461	90.1	90
86	6.2906	1701	0.4398	44	1244	0.9	119705	91.1	91
87	6.4806	1720	0.4326	43	2193	1.7	121898	92.7	92
88	6.6658	1738	0.4287	43	3088	2.3	124986	95.1	94
89	6.8491	1756	0.428	43	2854	2.2	127840	97.2	96
90	7.0331	1775	0.4305	43	776	0.6	128616	97.8	98
91	7.2208	1794	0.4365	44	0	0	128616	97.8	98
92	7.4154	1813	0.4464	45	3	0	128619	97.8	98
93	7.6209	1834	0.4611	46	5	0	128624	97.8	98
94	7.8427	1856	0.4821	48	21	0	128645	97.9	98
95	8.089	1880	0.5121	51	37	0	128682	97.9	98
96	8.3727	1909	0.5559	56	105	0.1	128787	98	98
97	8.718	1943	0.6241	62	274	0.2	129061	98.2	98
98	9.1784	1989	0.7441	74	554	0.4	129615	98.6	98
99	9.924	2064	1.0253	103	923	0.7	130538	99.3	99
100	11.1632	2188	1.8393	184	925	0.7	131463	100	99

WRITING GRADE 8

Raw Score	Measure	Scale Score	Logit SE	Scale Score SE	Freq	Freq %	Cum Freq	Cum Freq %	Percentile
22	-8.3505	700	1.8378	184	12	0	12	0	1
23	-7.1151	700	1.0225	102	74	0.1	86	0.1	1
24	-6.3755	700	0.7399	74	170	0.1	256	0.2	1
25	-5.9214	700	0.6188	62	249	0.2	505	0.4	1
26	-5.583	700	0.5495	55	264	0.2	769	0.5	1
27	-5.3067	700	0.5044	50	237	0.2	1006	0.7	1
28	-5.0685	700	0.4731	47	185	0.1	1191	0.8	1
29	-4.8557	700	0.4506	45	134	0.1	1325	0.9	1
30	-4.6604	700	0.4341	43	75	0.1	1400	1	1
31	-4.4774	700	0.4221	42	58	0	1458	1	1
32	-4.303	700	0.4136	41	48	0	1506	1.1	1
33	-4.1345	710	0.4078	41	50	0	1556	1.1	1
34	-3.9697	727	0.4043	40	140	0.1	1696	1.2	1
35	-3.8069	743	0.4029	40	270	0.2	1966	1.4	1
36	-3.6446	759	0.403	40	353	0.2	2319	1.6	1
37	-3.4817	776	0.4046	40	456	0.3	2775	1.9	2
38	-3.317	792	0.4073	41	472	0.3	3247	2.3	2
39	-3.1497	809	0.4107	41	443	0.3	3690	2.6	2
40	-2.9795	826	0.4144	41	353	0.2	4043	2.8	3
41	-2.8063	843	0.4178	42	285	0.2	4328	3	3
42	-2.6307	861	0.4203	42	253	0.2	4581	3.2	3
43	-2.4535	878	0.4212	42	175	0.1	4756	3.3	3
44	-2.2763	896	0.4203	42	168	0.1	4924	3.4	3
45	-2.1009	914	0.4172	42	283	0.2	5207	3.6	4
46	-1.9288	931	0.4122	41	543	0.4	5750	4	4
47	-1.7614	948	0.4057	41	1031	0.7	6781	4.7	4
48	-1.5999	964	0.3982	40	1441	1	8222	5.8	5
49	-1.4444	979	0.3904	39	2050	1.4	10272	7.2	6
50	-1.295	994	0.3826	38	2295	1.6	12567	8.8	8
51	-1.1514	1009	0.3752	38	2432	1.7	14999	10.5	10
52	-1.0132	1023	0.3684	37	2402	1.7	17401	12.2	11
53	-0.8797	1036	0.3623	36	2149	1.5	19550	13.7	13
54	-0.7504	1049	0.357	36	1907	1.3	21457	15	14
55	-0.6245	1061	0.3526	35	1497	1	22954	16.1	16
56	-0.5016	1074	0.3489	35	1034	0.7	23988	16.8	16
57	-0.3808	1086	0.3461	35	729	0.5	24717	17.3	17
58	-0.2618	1098	0.3441	34	691	0.5	25408	17.8	18
59	-0.1438	1109	0.343	34	950	0.7	26358	18.4	18
60	-0.0263	1121	0.3427	34	1481	1	27839	19.5	19
61	0.0913	1133	0.3433	34	2120	1.5	29959	21	20
62	0.2097	1145	0.3448	34	2697	1.9	32656	22.8	22
63	0.3293	1157	0.3472	35	3152	2.2	35808	25	24
64	0.4511	1169	0.3507	35	3396	2.4	39204	27.4	26
65	0.5756	1181	0.3553	36	3439	2.4	42643	29.8	29
66	0.7038	1194	0.3611	36	2958	2.1	45601	31.9	31
67	0.8367	1208	0.3683	37	2166	1.5	47767	33.4	33
68	0.9755	1221	0.3771	38	1107	0.8	48874	34.2	34
69	1.1216	1236	0.3877	39	613	0.4	49487	34.6	34
70	1.2767	1252	0.4004	40	1010	0.7	50497	35.3	35
71	1.4431	1268	0.4157	42	1869	1.3	52366	36.6	36
72	1.6234	1286	0.4339	43	3118	2.2	55484	38.8	38
73	1.8208	1306	0.4554	46	5090	3.6	60574	42.4	41
74	2.0395	1328	0.4803	48	7586	5.3	68160	47.7	45
75	2.2835	1352	0.5078	51	10638	7.4	78798	55.1	51
76	2.5559	1379	0.5355	54	12814	9	91612	64.1	60
77	2.8555	1409	0.5575	56	12673	8.9	104285	72.9	69
78	3.1728	1441	0.5664	57	8535	6	112820	78.9	76
79	3.4907	1473	0.5586	56	2110	1.5	114930	80.4	80
80	3.7924	1503	0.5385	54	286	0.2	115216	80.6	80
81	4.0693	1531	0.5138	51	86	0.1	115302	80.7	81
82	4.321	1556	0.4899	49	185	0.1	115487	80.8	81
83	4.5508	1579	0.4693	47	380	0.3	115867	81	81
84	4.7631	1600	0.4527	45	885	0.6	116752	81.7	81
85	4.962	1620	0.4398	44	1637	1.1	118389	82.8	82
86	5.1511	1639	0.4305	43	2781	1.9	121170	84.8	84
87	5.3335	1657	0.4244	42	4208	2.9	125378	87.7	86
88	5.5121	1675	0.4213	42	4968	3.5	130346	91.2	89
89	5.6894	1693	0.4212	42	3806	2.7	134152	93.8	93
90	5.8679	1711	0.4243	42	923	0.6	135075	94.5	94
91	6.0503	1729	0.4306	43	2	0	135077	94.5	94
92	6.2399	1748	0.4409	44	4	0	135081	94.5	94
93	6.4407	1768	0.456	46	40	0	135121	94.5	95
94	6.6579	1790	0.4774	48	82	0.1	135203	94.6	95
95	6.8997	1814	0.5077	51	233	0.2	135436	94.7	95
96	7.1791	1842	0.5521	55	472	0.3	135908	95.1	95
97	7.5202	1876	0.6207	62	971	0.7	136879	95.7	95
98	7.9763	1921	0.7413	74	1675	1.2	138554	96.9	96
99	8.7178	1996	1.0233	102	2353	1.6	140907	98.6	98
100	9.9544	2119	1.8382	184	2053	1.4	142960	100	99

WRITING GRADE 11

Raw Score	Measure	Scale Score	Logit SE	Scale Score SE	Freq	Freq %	Cum Freq	Cum Freq %	Percentile
22	-7.5521	700	1.8376	184	6	0	6	0	1
23	-6.3173	700	1.022	102	47	0	53	0	1
24	-5.5789	700	0.739	74	101	0.1	154	0.1	1
25	-5.1265	732	0.6174	62	180	0.1	334	0.3	1
26	-4.79	765	0.5474	55	213	0.2	547	0.4	1
27	-4.5162	793	0.5016	50	216	0.2	763	0.6	1
28	-4.2813	816	0.4694	47	163	0.1	926	0.7	1
29	-4.0724	837	0.4458	45	125	0.1	1051	0.8	1
30	-3.8818	856	0.4281	43	89	0.1	1140	0.9	1
31	-3.7045	874	0.4145	41	62	0	1202	0.9	1
32	-3.5372	891	0.4042	40	47	0	1249	0.9	1
33	-3.3771	907	0.3963	40	42	0	1291	1	1
34	-3.2225	922	0.3904	39	107	0.1	1398	1.1	1
35	-3.0718	937	0.3861	39	192	0.1	1590	1.2	1
36	-2.9239	952	0.3831	38	305	0.2	1895	1.4	1
37	-2.778	967	0.381	38	399	0.3	2294	1.7	2
38	-2.6334	981	0.3798	38	424	0.3	2718	2.1	2
39	-2.4895	995	0.379	38	454	0.3	3172	2.4	2
40	-2.3459	1010	0.3787	38	409	0.3	3581	2.7	3
41	-2.2027	1024	0.3784	38	331	0.3	3912	3	3
42	-2.0596	1038	0.3781	38	96	0.1	4008	3	3
43	-1.9167	1053	0.3776	38	95	0.1	4103	3.1	3
44	-1.7745	1067	0.3767	38	98	0.1	4201	3.2	3
45	-1.6331	1081	0.3753	38	156	0.1	4357	3.3	3
46	-1.4928	1095	0.3735	37	309	0.2	4666	3.5	3
47	-1.3542	1109	0.3711	37	637	0.5	5303	4	4
48	-1.2175	1123	0.3684	37	1097	0.8	6400	4.8	4
49	-1.0829	1136	0.3654	37	1416	1.1	7816	5.9	5
50	-0.9505	1149	0.3622	36	1741	1.3	9557	7.2	7
51	-0.8205	1162	0.3589	36	1790	1.4	11347	8.6	8
52	-0.6929	1175	0.3557	36	1939	1.5	13286	10	9
53	-0.5674	1188	0.3528	35	1923	1.5	15209	11.5	11
54	-0.4439	1200	0.35	35	1621	1.2	16830	12.7	12
55	-0.3223	1212	0.3477	35	1348	1	18178	13.7	13
56	-0.2021	1224	0.3458	35	916	0.7	19094	14.4	14
57	-0.083	1236	0.3444	34	790	0.6	19884	15	15
58	0.0353	1248	0.3435	34	598	0.5	20482	15.5	15
59	0.1532	1260	0.3433	34	887	0.7	21369	16.2	16
60	0.2712	1271	0.3437	34	1310	1	22679	17.2	17
61	0.3897	1283	0.3449	34	1795	1.4	24474	18.5	18
62	0.5092	1295	0.3468	35	2371	1.8	26845	20.3	19
63	0.6304	1307	0.3497	35	2991	2.3	29836	22.6	21
64	0.754	1320	0.3535	35	3393	2.6	33229	25.1	24
65	0.8806	1332	0.3584	36	3594	2.7	36823	27.9	26
66	1.0113	1345	0.3647	36	3551	2.7	40374	30.5	29
67	1.147	1359	0.3724	37	2820	2.1	43194	32.7	32
68	1.2891	1373	0.3819	38	1369	1	44563	33.7	33
69	1.4393	1388	0.3935	39	444	0.3	45007	34	34
70	1.5996	1404	0.4077	41	684	0.5	45691	34.6	34
71	1.7728	1422	0.4252	43	1305	1	46996	35.5	35
72	1.9627	1441	0.447	45	2196	1.7	49192	37.2	36
73	2.1744	1462	0.4742	47	3640	2.8	52832	40	39
74	2.4153	1486	0.5086	51	5851	4.4	58683	44.4	42
75	2.6957	1514	0.5517	55	8928	6.8	67611	51.1	48
76	3.0287	1547	0.6033	60	12726	9.6	80337	60.8	56
77	3.4255	1587	0.6545	65	15768	11.9	96105	72.7	67
78	3.875	1632	0.6792	68	12612	9.5	108717	82.2	77
79	4.3265	1677	0.6574	66	2410	1.8	111127	84.1	83
80	4.7289	1717	0.6093	61	238	0.2	111365	84.2	84
81	5.0709	1751	0.5611	56	30	0	111395	84.3	84
82	5.3633	1781	0.5218	52	40	0	111435	84.3	84
83	5.6195	1806	0.4916	49	138	0.1	111573	84.4	84
84	5.8496	1829	0.4689	47	287	0.2	111860	84.6	84
85	6.0614	1850	0.4523	45	741	0.6	112601	85.2	85
86	6.2602	1870	0.4404	44	1530	1.2	114131	86.3	86
87	6.4505	1889	0.4325	43	2862	2.2	116993	88.5	87
88	6.6355	1908	0.4283	43	4703	3.6	121696	92	90
89	6.8182	1926	0.4273	43	4298	3.3	125994	95.3	94
90	7.0016	1944	0.4296	43	890	0.7	126884	96	96
91	7.1884	1963	0.4354	44	0	0	126884	96	96
92	7.382	1983	0.4453	45	4	0	126888	96	96
93	7.5864	2003	0.4599	46	8	0	126896	96	96
94	7.8072	2025	0.481	48	16	0	126912	96	96
95	8.0524	2050	0.511	51	52	0	126964	96	96
96	8.335	2078	0.5549	55	133	0.1	127097	96.1	96
97	8.6791	2112	0.6232	62	411	0.3	127508	96.4	96
98	9.1383	2158	0.7433	74	962	0.7	128470	97.2	97
99	9.8827	2233	1.0248	102	1751	1.3	130221	98.5	98
100	11.1211	2356	1.8391	184	1989	1.5	132210	100	99

Appendix O:

Performance Level Descriptors for Writing

Pennsylvania Department of Education
Grade 5 Writing
Performance Level Descriptors

A student scoring at the **Advanced Level** produces narrative, informational, and persuasive pieces of writing that demonstrate a comprehensive command of composition skills. A student writing at this level

1. writes with a sharp, distinct focus that identifies topic and task
2. shows a sophisticated awareness of audience and mode
3. –
4. gathers, organizes, and selects substantial, effective content appropriate for topic, task, and audience
5. –
6. develops paragraphs with strong topic sentences and illustrative supporting details
7. crafts effective introductions, bodies, and conclusions
8. uses logical organizational structures and strategies within sentences and between paragraphs to thoroughly develop content
9. uses a variety of effective transitions to develop a controlling idea
10. varies lengths and patterns of simple and compound sentences
11. utilizes vivid and precise language to develop and maintain a consistent voice
12. revises writing to effectively improve organization, word choice, logic, order of ideas, and precision of vocabulary
13. demonstrates skill in editing to eliminate most errors in spelling, capitalization, punctuation, usage, and sentence structure

A student scoring at the **Proficient Level** produces narrative, informational, and persuasive pieces of writing that demonstrate a thorough understanding of composition skills. A student writing at this level

1. writes with a clear focus that identifies topic and task
2. shows a general awareness of audience and mode
3. –
4. gathers, organizes, and selects content appropriate for topic, task, and audience
5. –
6. develops paragraphs with topic sentences and relevant supporting details
7. produces adequate introductions, bodies, and conclusions
8. uses logical organizational structures and strategies within sentences and between paragraphs to sufficiently develop content
9. uses functional transitions to develop a controlling idea
10. varies lengths and patterns of simple and compound sentences
11. utilizes precise language to develop and maintain a consistent voice
12. revises writing to sufficiently address organization, word choice, logic, order of ideas, and precision of vocabulary
13. demonstrates skill in editing to eliminate common errors in spelling, capitalization, punctuation, usage, and sentence structure

A student scoring at the **Basic Level** produces narrative, informational, and persuasive pieces of writing that demonstrate a limited understanding of composition skills. A student writing at this level

1. writes with a vague or indistinct focus to identify topic and/or task
2. shows a limited awareness of audience and mode
3. –
4. needs assistance to gather and select content appropriate for topic, task, and audience
5. –
6. constructs under-developed paragraphs with unclear topic sentences and/or insufficient supporting details
7. produces inadequate introductions, bodies, and/or conclusions
8. shows limited ability to use logical organizational structures and/or strategies within sentences and/or between paragraphs to develop content
9. uses few and/or ineffective transitions
10. lacks variety in lengths and patterns of simple and compound sentences
11. utilizes vague or imprecise language often leading to an ineffective voice
12. demonstrates limited ability to revise writing
13. shows a limited ability to eliminate errors in spelling, capitalization, punctuation, usage, and sentence structure

A student scoring at the **Below Basic Level** produces writing that demonstrates a below grade-level understanding of composition skills and requires extensive assistance with composing, revising, and editing.

Pennsylvania Department of Education
Grade 8 Writing
Performance Level Descriptors

A student scoring at the **Advanced Level** produces narrative, informational, and persuasive pieces of writing that demonstrate a comprehensive command of composition skills. A student writing at this level

1. writes with a sharp, distinct focus that identifies topic and task
2. shows a sophisticated awareness of audience and mode
3. establishes a single point of view when appropriate
4. gathers valid and reliable information and organizes substantial, effective content appropriate for topic
5. employs most effective format for purpose and audience
6. develops paragraphs with illustrative supporting details specific to the topic and relevant to the focus
7. crafts effective introductions that establish topic and purpose; crafts effective conclusions that reiterate topic and purpose
8. uses logical and sophisticated organizational structures and strategies within sentences and between paragraphs to thoroughly develop content
9. uses a variety of effective transitions to develop a controlling idea
10. varies lengths and patterns of simple, compound, and complex sentences
11. utilizes vivid and precise language to maintain a consistent voice and tone
12. revises writing to effectively improve logic and organization, content, paragraph development, detail, style, tone, and word choice
13. demonstrates skill in editing to eliminate most errors in spelling, capitalization, punctuation, usage, and sentence structure

A student scoring at the **Proficient Level** produces narrative, informational, and persuasive pieces of writing that demonstrate a thorough understanding of composition skills. A student writing at this level

1. writes with a clear focus that identifies topic and task
2. shows a general awareness of audience and mode
3. establishes a single point of view when appropriate
4. gathers valid and/or reliable information and organizes content appropriate for topic
5. employs effective format for purpose and audience
6. develops paragraphs with supporting relevant details specific to the topic and relevant to the focus
7. produces adequate introductions that establish topic and purpose; produces adequate conclusions that reiterate topic and purpose
8. uses logical organizational structures and strategies within sentences and between paragraphs to sufficiently develop content
9. uses functional transitions to develop a controlling idea
10. varies lengths and patterns of simple, compound, and complex sentences
11. utilizes precise language to maintain a consistent voice and tone
12. revises writing after rethinking to sufficiently address logic and organization, content, paragraph development, detail, style, tone, and word choice

13. demonstrates skill in editing to eliminate common errors in spelling, capitalization, punctuation, usage, and sentence structure

A student scoring at the **Basic Level** produces narrative, informational, and persuasive pieces of writing that demonstrate a limited understanding of composition skills. A student writing at this level

1. writes with a vague or indistinct focus to identify topic and/or task
2. shows a limited awareness of audience and mode
3. may not establish a single point of view
4. needs assistance to gather valid and/or reliable information and organize content appropriate for topic
5. may employ ineffective format for purpose and/or audience
6. constructs under-developed paragraphs with insufficient supporting details
7. produces inadequate introductions and/or conclusions
8. shows limited ability to use logical organizational strategies within sentences and/or between paragraphs
9. uses few and/ineffective transitions
10. lacks variety in lengths and patterns of simple, compound, and/or complex sentences
11. utilizes vague or imprecise language often leading to an ineffective voice and/or tone
12. demonstrates limited ability to revise writing
13. shows a limited ability to eliminate errors in spelling, capitalization, punctuation, usage, and sentence structure

A student scoring at the **Below Basic Level** produces writing that demonstrates a below grade-level understanding of composition skills and requires extensive assistance with composing, revising, and editing.

Pennsylvania Department of Education
Grade 11 Writing
Performance Level Descriptors

A student scoring at the **Advanced Level** produces narrative, informational, and persuasive pieces of writing that demonstrate a comprehensive command of composition skills. A student writing at this level

1. writes with a sharp, distinct focus that identifies topic and task
2. shows a sophisticated awareness of audience and mode
3. establishes and maintains a single point of view when appropriate
4. gathers and organizes valid and reliable information; analyzes substantial, effective content appropriate for topic
5. employs most effective format for purpose and audience
6. writes fully-developed paragraphs with illustrative supporting details specific to the topic and relevant to the focus
7. crafts effective introductions and conclusions
8. uses logical and sophisticated organizational structures and strategies to thoroughly develop content
9. uses a variety of effective transitions to develop a controlling idea
10. varies lengths, types, and patterns of sentences
11. utilizes vivid and precise language throughout to maintain a consistent voice and tone
12. revises writing to effectively improve style, word choice, sentence variety and subtlety of meaning after rethinking purpose, audience, and genre
13. demonstrates skill in editing to eliminate most errors in spelling, capitalization, punctuation, usage, and sentence structure

A student scoring at the **Proficient Level** produces narrative, informational, and persuasive pieces of writing that demonstrate a thorough understanding of composition skills. A student writing at this level

1. writes with a clear focus that identifies topic and task
2. shows a general awareness of audience and mode
3. establishes and maintains a single point of view when appropriate
4. gathers and organizes valid and/or reliable information; analyzes content appropriate for topic
5. employs effective format for purpose and audience
6. writes well-developed paragraphs with relevant supporting details specific to the topic and relevant to the focus
7. produces adequate introductions and conclusions
8. uses logical organizational structures and strategies to sufficiently develop content
9. uses functional transitions to develop a controlling idea
10. varies lengths, types, and patterns of sentences
11. utilizes precise language to maintain a consistent voice and tone
12. revises writing to sufficiently address style, word choice, sentence variety and subtlety of meaning after rethinking purpose, audience, and genre

13. demonstrates skill in editing to eliminate common errors in spelling, capitalization, punctuation, usage, and sentence structure

A student scoring at the **Basic Level** produces narrative, informational, and persuasive pieces of writing that demonstrate a limited understanding of composition skills. A student writing at this level

1. writes with a vague or indistinct focus to identify topic and/or task
2. shows a limited awareness of audience and mode
3. may establish but not maintain a single point of view
4. needs assistance to gather valid and/or reliable information and organize content appropriate for topic
5. may employ ineffective format for purpose and/or audience
6. constructs under-developed paragraphs with insufficient supporting details
7. produces inadequate introductions and/or conclusions
8. shows a limited ability to use logical organizational structures and/or strategies to develop content
9. uses few and/or ineffective transitions
10. lacks variety in types and patterns of sentences
11. utilizes vague or imprecise language often leading to an ineffective voice and/or tone
12. demonstrates limited ability to revise writing
13. shows a limited ability to eliminate errors in spelling, capitalization, punctuation, usage, and sentence structure

A student scoring at the **Below Basic Level** produces writing that demonstrates a below grade-level understanding of composition skills and requires extensive assistance with composing, revising, and editing.

Appendix P:

PSSA Writing Standards Setting Meeting Agenda

Pennsylvania Writing Standards Setting Meeting
June 20 – 21, 2006
Grantville, Pennsylvania

Tuesday, June 20, 2006

7:30 am – 8:00 am	Breakfast – Restaurant
8:00 am – 8:30 am	Check-in – *Grande Foyer*
8:30 am – 9:00 am	Introduction to Standards Setting. Shula Nedley, Ray Young – PDE introduction. David Chayer – DRC introduction. John Born – Reimbursement and other administrative procedures. – *Grande II*
9:00 am – 9:30 am	Overview of Method
9:30 am – 10:15 am	Training using sample materials.
10:15 am – 10:30 am	Morning Break – Participants move to breakout rooms by grade levels (grade 5 – *Grande II*, grade 8 – *Grande III* or grade 11 – *Grande IV*)
10:30 am – 11:00 am	Take the Operational Test
11:00 am – 11:30 am	Presentation and Group Discussion of Performance Level Descriptors
11:30 am – 12:30 pm	Lunch – Restaurant
12:30 pm – 3:30 pm	Range Finding
3:30 pm – 3:45 pm	Afternoon Break
3:45 pm – 5:00 pm	Group Discussion and Revisions
5:00 pm	Check-in Materials

Wednesday, June 21, 2006

7:30 am – 8:00 am	Breakfast – Restaurant
8:00 am – 8:30 am	Check-in – *Grande Foyer*
8:30 am – 9:00 pm	Group Discussion of Impacts (grade 5 – *Grande II*, grade 8 – *Grande III* or grade 11 – *Grande IV*)
9:00 am – 10:15 am	Pinpointing
10:15 am – 10:30 am	Morning Break
10:30 am – 12:00 pm	Pinpointing Continued
12:00 pm – 1:00 pm	Lunch – Restaurant
1:00 pm – 3:00 pm	Pinpointing Continued
3:00 pm – 3:15 pm	Afternoon Break
3:15 pm – 4:30 pm	Group Discussion and Revisions
4:30 pm	Check-in Materials

Appendix Q:

PSSA Writing Standards Setting Meeting Participants' Evaluation Results Summary

Pennsylvania Standards Setting
June 20-21, 2006

Grade 5 Writing Evaluation Form

The purpose of this Evaluation Form is to obtain your opinions about the standards setting. Your opinions will provide a basis for evaluating both the materials and the training. We request that you **not** put your name on this form. We want your opinions to remain anonymous.

1. Check the column that most accurately reflects your opinion regarding the usefulness of the following materials used in the standards setting: **N=10**

Materials	Not Useful	Partially Useful	Useful	Very Useful
Performance Level Descriptors	0%	0%	0%	**100%**
Item Map	0%	0%	0%	**100%**
Items	0%	0%	10%	**90%**
Samples of Student Responses	0%	0%	0%	**100%**
Rubrics	0%	0%	10%	**90%**

2. Indicate the importance of the following factors in your classifications: **N=10**

Factor	Not Important	Somewhat Important	Important	Very Important
Descriptions of Below Basic, Basic, Proficient and Advanced	0%	0%	0%	**100%**
Your own classroom experience	0%	10%	20%	**70%**
Multiple Choice Items	0%	20%	20%	**60%**
Writing Prompt Responses	0%	0%	0%	**100%**
Group discussions	0%	0%	0%	**100%**
Other panelists' placements into the four performance levels	0%	0%	**70%**	30%

3. Check the column that reflects your confidence in your final judgment for the four performance levels: **N=10**

Performance Level	Not Confident	Partially Confident	Confident	Very Confident
Below Basic/Basic	0%	0%	40%	**60%**
Basic/Proficient	0%	10%	30%	**60%**
Proficient/ Advanced	0%	0%	30%	**70%**

3. How adequate was the training provided on the tasks to prepare you for your subsequent judgments? **N=10**

a. Not Adequate 0%
b. Partially Adequate 0%
c. Adequate 80%
d. Very Adequate 20%

4. How would you rate the amount of time used for training? **N=10**

a. Too little time 30%
b. About right 70%
c. Too much time 0%

5. How would you rate the amount of time allotted for your judgements **after** the training? **N=10**

a. Too little time 10%
b. About right 70%
c. Too much time 20%

6. How confident are you that the processes and methods used for the standards setting will produce a reliable and valid result? **N=10**

a. Not Confident 0%
b. Partially Confident 0%
c. Confident 80%
d. Very Confident 20%

7. How would you rate the facilities? **N=10**

a. Not Suitable 0%
b. Somewhat Suitable 20%
c. Highly Suitable 80%

Please provide us with your suggestions for ways to improve the standards setting in the space provided (if you need additional space, you may continue on the back of this page):

We should spread out more across tables.

The table leader provided no input during this process.

Proper disclosure of the training process needs to take place.

Pennsylvania Standards Setting
June 20-21, 2006

Grade 8 Writing Evaluation Form

The purpose of this Evaluation Form is to obtain your opinions about the standards setting. Your opinions will provide a basis for evaluating both the materials and the training. We request that you **not** put your name on this form. We want your opinions to remain anonymous.

1. Check the column that most accurately reflects your opinion regarding the usefulness of the following materials used in the standards setting: **N=10**

Materials	Not Useful	Partially Useful	Useful	Very Useful
Performance Level Descriptors	0%	0%	10%	**90%**
Item Map	0%	0%	10%	**90%**
Items	0%	20%	10%	**70%**
Samples of Student Responses	0%	0%	0%	**100%**
Rubrics	0%	10%	10%	**80%**

2. Indicate the importance of the following factors in your classifications: **N=10**

Factor	Not Important	Somewhat Important	Important	Very Important
Descriptions of Below Basic, Basic, Proficient and Advanced	0%	0%	0%	**100%**
Your own classroom experience	0%	10%	20%	**70%**
Multiple Choice Items	0%	20%	**40%**	**40%**
Writing Prompt Responses	0%	0%	0%	**100%**
Group discussions	0%	0%	40%	**60%**
Other panelists' placements into the four performance levels	0%	20%	20%	**60%**

3. Check the column that reflects your confidence in your final judgment for the four performance levels: **N=10**

Performance Level	Not Confident	Partially Confident	Confident	Very Confident
Below Basic/Basic	0%	20%	**40%**	**40%**
Basic/Proficient	0%	10%	30%	**60%**
Proficient/ Advanced	0%	10%	40%	**50%**

3. How adequate was the training provided on the tasks to prepare you for your subsequent judgments? **N=10**

a. Not Adequate 0%
b. Partially Adequate 0%
c. Adequate 40%
d. Very Adequate 60%

4. How would you rate the amount of time used for training? **N=10**

a. Too little time 0%
b. About right 100%
c. Too much time 0%

5. How would you rate the amount of time allotted for your judgements **after** the training? **N=10**

a. Too little time 10%
b. About right 90%
c. Too much time 0%

6. How confident are you that the processes and methods used for the standards setting will produce a reliable and valid result? **N=10**

a. Not Confident 0%
b. Partially Confident 20%
c. Confident 40%
d. Very Confident 40%

7. How would you rate the facilities? **N=10**

a. Not Suitable 0%
b. Somewhat Suitable 20%
c. Highly Suitable 80%

Please provide us with your suggestions for ways to improve the standards setting in the space provided (if you need additional space, you may continue on the back of this page):

Pennsylvania Standards Setting
June 20-21, 2006

Grade 11 Writing Evaluation Form

The purpose of this Evaluation Form is to obtain your opinions about the standards setting. Your opinions will provide a basis for evaluating both the materials and the training. We request that you **not** put your name on this form. We want your opinions to remain anonymous.

1. Check the column that most accurately reflects your opinion regarding the usefulness of the following materials used in the standards setting: **N=7**

Materials	Not Useful	Partially Useful	Useful	Very Useful
Performance Level Descriptors	0%	0%	28.6%	**71.4%**
Item Map	0%	14.3%	28.6%	**57.1%**
Items	0%	14.3%	0%	**85.7%**
Samples of Student Responses	0%	0%	14.3%	**85.7%**
Rubrics	14.3%	28.6%	0%	**57.1%**

2. Indicate the importance of the following factors in your classifications: **N=7**

Factor	Not Important	Somewhat Important	Important	Very Important
Descriptions of Below Basic, Basic, Proficient and Advanced	0%	0%	14.3%	**85.7%**
Your own classroom experience	0%	14.3%	14.3%	**71.4%**
Multiple Choice Items	14.3%	14.3%	**42.9%**	28.6%
Writing Prompt Responses	0%	0%	14.3%	**85.7%**
Group discussions	0%	14.3%	28.6%	**57.1%**
Other panelists' placements into the four performance levels	0%	14.3%	**42.9%**	**42.9%**

3. Check the column that reflects your confidence in your final judgment for the four performance levels: **N=6**

Performance Level	Not Confident	Partially Confident	Confident	Very Confident
Below Basic/Basic	0%	0%	33.3%	**66.7%**
Basic/Proficient	0%	0%	0%	**100%**
Proficient/ Advanced	0%	0%	16.7%	**83.3%**

3. How adequate was the training provided on the tasks to prepare you for your subsequent judgments? **N=6**

a. Not Adequate 0%
b. Partially Adequate 0%
c. Adequate 50%
d. Very Adequate 50%

4. How would you rate the amount of time used for training? **N=6**

a. Too little time 16.7%
b. About right 83.3%
c. Too much time 0%

5. How would you rate the amount of time allotted for your judgements **after** the training? **N=6**

a. Too little time 0%
b. About right 100%
c. Too much time 0%

6. How confident are you that the processes and methods used for the standards setting will produce a reliable and valid result? **N=6**

a. Not Confident 0%
b. Partially Confident 0%
c. Confident 50%
d. Very Confident 50%

7. How would you rate the facilities? **N=6**

a. Not Suitable 0%
b. Somewhat Suitable 0%
c. Highly Suitable 100%

Please provide us with your suggestions for ways to improve the standards setting in the space provided (if you need additional space, you may continue on the back of this page):