PREFACE: An Overview of Recent and Future Assessments

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 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 *2005 PSSA Technical Report for 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, standards 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 *2006 PSSA Technical Report for 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. In 2007 the grade 3 reading and mathematics assessment became DRC's responsibility and is covered in the present technical report, along with grades 4 through 8, and 11.

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. 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, standards setting, and other technical characteristics of the operational 2006 PSSA may be found in the *2006 PSSA Technical Report for Writing*.

The introduction of an operational science assessment in 2008 moved closer to reality with a major standalone field test at grades 4, 8, and 11 in April–May of 2007. A description of the development of science scenarios and related multiple-choice, short answer open-ended, and extended open-ended questions, item review processes, statistical analysis of field test data, and selection of items for the 2008 operational science test may be found in the *2008 PSSA Preliminary Technical Report for Science*.

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).

ASSESSMENT ACTIVITIES OCCURRING IN THE 2003–04 SCHOOL YEAR

Table P–1 outlines the operational assessments and field tests 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 *2005 PSSA Technical Report for 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 2004
5	Operational mathematics and reading	April 2004
	Standalone field test in mathematics and reading	April/May 2004
6	Operational writing	October 2004
8	Operational mathematics and reading	April 2004
	Standalone field test in mathematics and reading	April/May 2004
9	Operational writing	October 2004
11	Operational mathematics and reading	April 2004
	Standalone field test in mathematics and reading	April/May 2004
	Operational writing	February 2004
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 2004

ASSESSMENT ACTIVITIES OCCURRING IN THE 2004–05 SCHOOL YEAR

Table P–2 displays the operational assessments and field tests 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 assessment was reassessed 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 2005
4	Standalone field test for mathematics and reading	April 2005
5	Operational mathematics and reading with embedded field test	April 2005
	Standalone field test in writing	February 2005
6	Standalone field test for mathematics and reading	April 2005
7	Standalone field test for mathematics and reading	April 2005
8	Operational mathematics and reading with embedded field test	April 2005
	Standalone field test in writing	February 2005
11	Operational mathematics and reading with embedded field test	April 2005
	Operational writing with embedded field test	February 2005
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 2004

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 for operational assessments. 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 assessments 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 assessments featured mode-specific scoring guidelines; stimulus-based, multiple-choice items; and a grade-specific emphasis shift in writing modes assessed. See the *2006 PSSA Technical Report for Writing: Grades 5, 8, and 11* for further information about the new writing assessments. 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 2006
4	Operational mathematics and reading with embedded field test	March 2006
5	Operational mathematics and reading with embedded field test	March 2006
	Operational writing with embedded field test	February 2006
6	Operational mathematics and reading with embedded field test	March 2006
7	Operational mathematics and reading with embedded field test	March 2006
8	Operational mathematics and reading with embedded field test	March 2006
	Operational writing with embedded field test	February 2006
11	Operational mathematics and reading with embedded field test	March 2006
	Operational writing with embedded field test	February 2006
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 2005

ASSESSMENT ACTIVITIES OCCURRING IN THE 2006–07 SCHOOL YEAR

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

The operational assessment for writing at grades 5, 8, and 11 continued 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 needed to be field tested in 2007. However, new multiple-choice items were field tested in the 2007 writing assessment.

Following the spring operational assessments in writing and reading and mathematics, a separate, "standalone" field test in science occurred 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

Grade	Assessment Activity	Date
3	Operational mathematics and reading with embedded field test	March 2007
4	Operational mathematics and reading with embedded field test	March 2007
	Standalone field test in science	April/May 2007
5	Operational mathematics and reading with embedded field test	March 2007
	Operational writing with embedded field test	February 2007
6	Operational mathematics and reading with embedded field test	March 2007
7	Operational mathematics and reading with embedded field test	March 2007
8	Operational mathematics and reading with embedded field test	March 2007
	Operational writing with embedded field test	February 2007
	Standalone field test in science	April/May 2007
11	Operational mathematics and reading with embedded field test	March 2007
	Operational writing with embedded field test	February 2007
	Standalone field test in science	April/May 2007
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 2006

ASSESSMENT ACTIVITIES PLANNED FOR THE 2007–08 SCHOOL YEAR

Table P–5 shows the assessment plan for the 2007–08 school year. The mathematics and reading assessments will be operational for grades 3 through 8 and at grade 11. Field testing for mathematics and reading will continue to be embedded as part of the operational assessments at each grade level. As in the previous years, the retest opportunity at grade 12 continued.

The operational assessment for writing at grades 5, 8, and 11 continues in a February test window using mode-specific scoring guidelines; stimulus-based, multiple-choice items; and a grade-specific emphasis 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 will be field tested in 2008. However, new multiple-choice items will be field tested in the 2008 writing assessment.

The first operational assessment in science will be fully implemented in April/May. Similar to the other operational assessments, field testing for science will be embedded as part of the operational assessments at each grade level.

Table P–5. Operational Assessment and Field Testing During the 2007–08 School Year (Planned)

Grade	Assessment Activity	Date
3	Operational mathematics and reading with embedded field test	March/April 2008
4	Operational mathematics and reading with embedded field test	March/April 2008
	Operational science with embedded field test	April/May 2008
5	Operational mathematics and reading with embedded field test	March/April 2008
	Operational writing with embedded field test	February 2008
6	Operational mathematics and reading with embedded field test	March/April 2008
7	Operational mathematics and reading with embedded field test	March/April 2008
8	Operational mathematics and reading with embedded field test	March/April 2008
	Operational writing with embedded field test	February 2008
	Operational science with embedded field test	April/May 2008
11	Operational mathematics and reading with embedded field test	March/April 2008
	Operational writing with embedded field test	February 2008
	Operational science with embedded field test	April/May 2008
12	Retest opportunity for students who as grade 11 students in the spring of 2007 failed to reach at least the proficient level in mathematics, reading, writing, or science	October/ November 2007

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. The third year of calculations for Adequate Yearly Progress (AYP) were conducted and reported for grades 5, 8, and 11.

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. The fourth year of calculations for AYP were conducted and reported for grades 5, 8, and 11 and for the first time in grade 3.

In 2007 the operational reading and mathematics assessment continued in grades 3 through 8 and 11. AYP calculations for grades 4, 6, and 7 took place in 2007 when they were 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 the 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 the 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 assessments.

The 2006 PSSA operational writing assessment featured additional revisions 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.
- Stimulus-based revising/editing multiple-choice items were incorporated to provide a more reliable and valid measure of the conventions Academic Standard.

The 2007 PSSA operational writing assessment continued with the same structure and at the same time of year as in 2006.

THE PENNSYLVANIA SCIENCE ASSESSMENT

In accordance with the NCLB requirement to implement an operational science assessment in 2008, a major test development effort in science took place during 2006, followed by a large-scale, standalone field test in April/May of 2007. A full implementation of an operational science assessment at grades 4, 8, and 11 is scheduled for 2008.

Several historical milestones were significant to the development of a science test in Pennsylvania. These include:

- Adoption of Act 16 or Pennsylvania Senate Bill 652 in 2000, which redefined the PSSA "as a test developed and implemented by the Department of Education to determine only academic achievement relating directly to objective Academic Standards in the areas of reading, mathematics, and science." (see the *Science Assessment Handbook*, PDE, November 2006).
- Pennsylvania State Board of Education adoption of *Science and Technology Standards* on July 12, 2001 and the *Environment and Ecology Standards* on January 5, 2002.

Aligned to the *Pennsylvania Science Assessment Anchor Content Standards* and Eligible Content, the science test is designed to measure and report results in four major categories:

A. The Nature of Science,

B. Biological Sciences,

C. Physical Sciences, and

D. Earth and Space Sciences.

At grade 4, test questions consist of standalone multiple-choice and two-point short answer open-ended items, and at grade 8 and 11 test questions also consist of sets of multiple-choice questions associated with science scenarios. Grade 11 also has four-point open-ended items associated with the science scenarios. A science scenario consists of a description of a class project, an experiment, or other research. Scenarios typically contain text, graphs, charts and/or tables. Students use their content knowledge and science process skills to answer a set of multiple-choice items and, at grade 11 only, a four-point extended open-ended item related to the scenario. More information may be found in the following two Pennsylvania Department of Education publications available on the PDE website: *Science Assessment Handbook* and *2006–2007 Science Item and Scoring Sampler*.

An extensive description of the science test development activities, field testing, and statistical analyses may be found in the *2008 PSSA Preliminary Technical Report for Science*.

Chapter Two: New Test Development Required by NCLB

Spurred by PL 107-110, the *No Child Left Behind Act* of 2001 (NCLB), the Pennsylvania Department of Education (PDE) began to develop plans to expand testing into other grade levels and to design a standards-based assessment for science. Although grade 3 reading and mathematics tests were developed and administered statewide in 2003 and 2004, reporting results in terms of proficiency levels occurred for the first time in 2005. Reading and mathematics test development in the new grade levels of 4, 6, and 7 took place in 2004, with field testing occurring in 2005 and full implementation occurring in 2006. A field test for science occurred in 2007 with full implementation planned for 2008.

ASSESSMENT ANCHOR CONTENT STANDARDS AND ELIGIBLE CONTENT

The Academic Standards indicate what students should know and be able to do. Educator concerns regarding the number and breadth of Academic Standards led to an initiative by the Pennsylvania Department of Education (PDE) to develop Assessment Anchor Content Standards (Assessment Anchors) to indicate which parts of the Academic Standards (Instructional Standards) would be assessed on the PSSA. Based on recommendations from Pennsylvania educators, the Assessment Anchors (PDE, 2004) were designed as a tool to improve the articulation of curricular, instructional, and assessment practices. See Appendices A and B for examples of Assessment Anchor integration for mathematics and reading. They also serve to communicate Eligible Content, also called "assessment limits," or the range of knowledge and skills from which the PSSA would be designed.

A draft version of the Assessment Anchors and Eligible Content was submitted to Achieve, Inc., Washington, D.C., to conduct a special analysis to evaluate the degree of alignment with the Academic Standards. Preliminary feedback enabled PDE to make adjustments to improve the alignment as the Assessment Anchors took final form. These adjustments were reflected operationally starting with the 2007 PSSA.

Since the Assessment Anchors encompass grades 3 through 8 and grade 11, the document informs test design for all grade levels—distinguishing the Assessment Anchors in one grade from the Assessment Anchors in another grade and clarifying grade-level rigor.

OVERVIEW OF THE 2007 PSSA

The 2007 PSSA reading and mathematics tests contain items designed to reflect the revised Assessment Anchors. They were extensively reviewed and field tested in 2005 and 2006.

Mathematics Assessment Measures

The 2007 PSSA mathematics assessment has five major reporting categories: Numbers and Operations, Algebraic Concepts, Geometry, Measurement, and Data Analysis and Probability. By organizing the Assessment Anchors into a five-category reporting structure, there is a similarity to the categories used by the National Council of Teachers of Mathematics (NCTM) and the National Assessment of Educational Progress (NAEP).

The 2007 PSSA mathematics assessment employs two types of test items: multiple-choice and open-ended. These item types assess different levels of knowledge and provide different kinds of information about mathematics achievement. Psychometrically, multiple-choice items are very useful and efficient tools for collecting information about a student's academic achievement.

Open-ended performance tasks are less efficient in the sense that they generally generate fewer scorable points in the same amount of testing time. They do, however, provide tasks that are more realistic and better sample higher-level thinking skills. The design of the 2007 PSSA attempts to achieve a reasonable balance between the two item types. Furthermore, well-constructed scoring guides have made it possible to include open-ended tasks in large-scale assessments such as the PSSA. Trained scorers can apply the scoring guides to efficiently score large numbers of student papers in a highly reliable way.

MULTIPLE-CHOICE ITEMS

The majority of the mathematics items included on the 2007 PSSA are multiple-choice , or selected-response items. This item type is especially efficient for measuring a broad range of content. In the PSSA mathematics assessment, each multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Distractors typically represent incorrect concepts, incorrect logic, incorrect application of an algorithm, or computation errors.

Multiple-choice items are used to assess a variety of skill levels, from short-term recall of facts to problem solving. PSSA items involving application emphasize the requirement to carry out some mathematical process to find an answer, rather than simply recalling information from memory.

OPEN-ENDED TASKS FOR MATHEMATICS

Open-ended, or constructed-response tasks, require students to read a problem description and to develop an appropriate solution. The 2007 open-ended items are designed to take about ten minutes per task. Most of the open-ended items are designed to be scaffolded, which means that they have several components to the overall task that may enable students to enter or begin the problem at different places. In some items, each successive component is designed to assess progressively more difficult skills or higher knowledge levels. Certain components ask students to explain their reasoning for engaging in particular mathematical operations or for arriving at certain conclusions. The types of tasks utilized do not necessarily require computations. Students may also be asked to perform such tasks as constructing a graph, shading some portion of a figure, or listing object combinations that meet specified criteria.

Open-ended tasks are especially useful for measuring students' problem-solving skills in mathematics. They offer the opportunity to present real-life situations that require students to solve problems using math abilities learned in the classroom. Students must read the task carefully, identify the necessary information, devise a method of solution, perform the calculations, enter the solution directly in the answer document, and when required, offer an explanation. This provides insight into the students' mathematical knowledge, abilities, and reasoning processes.

The open-ended mathematics items are scored on a 0–4 point scale with an item-specific scoring guideline. The item-specific scoring guideline outlines the requirements at each score point. Item-specific scoring guidelines are based on the General Description of Mathematics Scoring Guidelines for Open-Ended Items. The general guidelines describe a hierarchy of responses, which represent the five score levels. See Appendix C or the grade-specific *Mathematics Item and Scoring Sampler*, available on the PDE website.

Reading Assessment Measures

The 2007 PSSA reading assessment has two major reporting categories, Comprehension and Reading Skills and Interpretation and Analysis of Fictional and Nonfictional Text. These two reporting categories are derived from Reading Academic Standards 1.1, 1.2, and 1.3. Standards 1.6, 1.7, and 1.8 are not addressed on the PSSA because they are not specific to reading comprehension and can be more accurately evaluated at the school level. Standards 1.4 and 1.5 are addressed on the PSSA writing assessment.

The reading assessment employs two types of test items: multiple-choice and open-ended. They are designed to measure students' comprehension of the information contained in the reading passages.

MULTIPLE-CHOICE ITEMS

Multiple-choice, or selected-response, items measure such concepts as how well students comprehend the overall meaning of a passage or make basic inferences about it. At times, asking students to choose a preferred answer is the best way to determine whether they have gleaned certain important information from a story. Such information may include setting, central idea, or main events and their sequence.

Each reading multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Incorrect response choices, or distractors, typically represent some kind of misinterpretation, predisposition, unsound reasoning, or casual reading.

OPEN-ENDED TASKS FOR READING

Open-ended, or constructed-response, tasks are designed to address comprehension of text in ways that multiple-choice items cannot. A short written response, requiring about ten minutes per item, allows students to prepare an answer and summarize using supporting details or examples derived from the text.

The reading open-ended items are scored on a 0–3 point scale with an item-specific scoring guideline. This scale is consistent with the scale used on the National Assessment of Educational Progress (NAEP). The change from the former 0–4 point scale improves the alignment with the types of tasks required. Each task is text-dependent and is carefully constructed with the scoring guide reflecting the task requirements. All item-specific scoring guidelines are based on the General Scoring Guidelines for Open-Ended Reading Items. The general guidelines describe a hierarchy of responses, which represent the four score levels. (See Appendix D or the grade-specific *Reading Item and Scoring Sampler*, available on the PDE website.)

Matrix Sampling Assessment Design

The PSSA was originally designed as a complex matrix-sampling scheme for both mathematics and reading, which was very efficient for measuring *school-level* performance, but less efficient for providing *student-level* assessments and diagnostics. In the present design, all forms contain a *common* core of items to which all students respond and *matrix* items that vary by form. Both the common and the matrix sections of the 2007 PSSA use traditional multiple-choice items and open-ended performance tasks. The forms are *spiraled* so that all forms are distributed uniformly within each testing room. This ensures that each matrix section is administered to an unbiased and sequentially random sample of students in each school. Since multiple forms are

administered, the blocks of matrix items expand the number of items available to more broadly measure the Assessment Anchors for school-level reporting.

The design changes that began to take effect with the spring 2000 administration shifted the measurement focus toward the student and away from the school. Beginning in 2000, student-level results were reported on an individual student report with diagnostic results at the *Academic Content Standard* level. All student-level results were based on the common items only and presented in the raw-score, percent-correct metric. In order to accommodate this change in focus, the common section was expanded to better reflect the curriculum. To administer the tests in a reasonable length of time, enhancing the common sections required a compensatory reduction of the matrix sections.

The PSSA design from 2000 through 2007 is an attempt to have the best of both worlds:

- All student-level results are based on the common core of items that all students in a grade are administered. This ensures that all students are evaluated using the same set of items.
- School-level content area total score results are based on the mean of the student-level scaled scores. This ensures that the results used for school accountability directly reflect the student-level results.
- School-level results at the Content Standard (Academic Standards category) level are based on the common items together with all embedded operational items on the matrix forms (embedded field-test items are not included in school-level results). This ensures that decisions regarding potential strengths and weaknesses at the school level better sample the entire curriculum.