Prairie State Achievement Examination

Technical Manual

2012 Testing Cycle

ACT and the Illinois State Board of Education

Table of Contents

List of Figures

List of Tables

Table Page

Preface

This manual documents the technical characteristics of the 2012 Prairie State Achievement Examination (PSAE) in light of its intended purposes. The PSAE is a two-day examination. Day 1 comprises the four tests of the ACT®. Day 2 comprises two WorkKeys® assessments (*Applied Mathematics* and *Reading for Information*) and an ISBE-developed science test.

Chapter 1 provides an overview of the PSAE. Chapter 2 provides evidence of validity of the PSAE in terms of the purposes for which the PSAE is to be used in Illinois. Chapter 3 provides evidence of the use of procedures and their results for sensitivity and bias reviews and DIF analysis. Chapter 4 shows documentation of the scaling process, reliability, measurement error, and generalizability of the PSAE for all content areas of the PSAE. Chapter 5 provides documentation of classification consistency for the PSAE. Chapter 6 documents the procedures for ensuring consistency of PSAE score meaning over time. Chapter 7 documents the quality control procedures for scoring, analysis, and reporting. Chapter 8 provides the results of the 2012 administration of the PSAE and Chapter 9 provides results for the 2012 PSAE Illinois State Goals Reports.

We encourage individuals who want more detailed information on topics that are discussed in this manual, or on related topics, to contact the Student Assessment Division of the Illinois State Board of Education.

Chapter 1
The Prairie State Achievement Examination

Overview and Purpose of the Prairie State Achievement Examination

The Illinois State Board of Education (ISBE) developed and adopted the Prairie State Achievement Examination (PSAE) in response to state and federal legislation. The federal Elementary and Secondary Education Act of 1994 requires states to (1) adopt challenging content and student performance standards and (2) demonstrate that they have adopted a set of high-quality yearly student assessments. In compliance with this law, ISBE adopted the Illinois Learning Standards in 1997. These standards are a set of statements that define the specific knowledge and skills that every public school student should learn in school. More than 28,000 Illinois citizens—including teachers, parents, school administrators, employers, community leaders, and representatives of higher education—participated in their development over a period of two years. The Illinois Learning Standards address student learning in seven areas: English language arts; mathematics; science; social science; physical development and health; fine arts; and foreign language.

To comply with the requirement for a high-quality, yearly student assessment at the high school level, the Illinois General Assembly established the PSAE through legislation passed on July 29, 1999 (Public Act 91-283). The PSAE is the regular statewide academic assessment that Illinois law requires public high school students to take. It is given to grade 11 students to measure their achievement with respect to the Illinois Learning Standards. The results of the PSAE may not be used as a graduation requirement that could prevent a student from receiving a high school diploma; however, legislation enacted in 2004 requires students to take the PSAE as a condition to receive a regular high school diploma, unless exempt.

Students took the PSAE for the first time in April 2001. In alignment with the Illinois Learning Standards and in accordance with current state law (105 ILCS 5/2-3.64), the 2012 PSAE assesses three academic subjects: reading, mathematics, and science.

Components of the PSAE

The PSAE comprises assessments from three sources: (1) the ACT®, which includes tests in English, mathematics, reading, and science; (2) an ISBE-developed science test; and (3) two WorkKeys® assessments (*Reading for Information* and *Applied Mathematics*). Table 1.1 shows how these components combine to produce the three PSAE subject tests.

Table 1.1: The Components of the PSAE

PSAE test scores		Component tests
Reading	→	ACT Reading Test + WorkKeys *Reading for Information*
Mathematics	→	ACT Mathematics Test + WorkKeys *Applied Mathematics*
Science	→	ACT Science Test + ISBE-developed science test

Purposes of the PSAE

The PSAE has three purposes: (1) to measure students' progress toward meeting the Illinois Learning Standards for state and federal accountability requirements, (2) to recognize the achievement of individual students who earn a Prairie State Achievement Award for excellent performance, and 3) to allow the receipt of a regular high school diploma by taking the test, unless exempt.

Population Served by the PSAE

All eligible grade 11 public-school students take the PSAE. In 2009, state legislation (Senate Bill 2014) eliminated the fall administration of the PSAE (the PSAE grade 12) that had been held in previous years.

Students with disabilities have the option of taking the PSAE under conditions that accommodate their individual disabilities. Students whose Individualized Education Programs (IEPs) identify the PSAE as being inappropriate for them, even with accommodations, are required to take the Illinois Alternate Assessment (IAA). All students with limited English proficiency (LEP) must take the PSAE. Students who have been in a state-approved Transitional Bilingual Education (TBE) program or Transitional Program of Instruction (TPI) for five or fewer years are eligible to test under State-Allowed Accommodations (see p. 3).

In April 2012, the PSAE was administered in Illinois in grade 11. Table 1.2 presents the demographic characteristics of the grade 11 students tested in 2012.

Table 1.2: Demographic Characteristics of Grade 11 Students Taking the Spring 2012 PSAE (Reported as Percentages)

Gender	Percent
Female	50
Male	50
No response	<1
Race/Ethnicity	
American Indian or Alaska Native	<1
Asian	4
Native Hawaiian or Other Pacific Islander	<1
Black or African American	18
Hispanic or Latino	20
White	55
Two or More Races	2
No response	<1

Administration of the PSAE

The PSAE is administered annually over a two-day period in April. Day 1 consists of the ACT and Day 2 consists of the ISBE-developed science test and the two WorkKeys assessments. Table 1.3 presents the April 2012 test-administration schedule for the PSAE. A makeup test (also given in a two-day period using the same schedule) is administered two weeks after the initial April test dates for students who miss one or both days of the initial administration.

It is critically important that the PSAE be administered under secure, standardized conditions. If a violation of certain administration conditions occurs during Day 1 testing (the ACT), scores could be voided or cancelled. Both self-reported and ACT-detected irregularities in the ACT test administration are reviewed at ACT, and may result in further investigation by ACT test compliance office staff. Under certain predetermined test administration conditions, scores will be reported for state reporting purposes only; that is, the scores may be used to calculate a student's PSAE score, but a college reportable ACT score will not be issued. Determinations of scoring eligibility for the PSAE are made in accordance with a scoring conditions document developed by ACT and approved by ISBE.

Training prior to test administration dates was required to ensure that newly appointed staff named as test supervisors, back-up test supervisors, or test accommodations coordinators were prepared to conduct a standardized test administration. Previously trained staff were encouraged, but not required, to participate in test administration training. In consideration of expense and time for all staff involved in the PSAE administration, all training was made available online in 2012 as a Webinar recording for appointed staff to view at their own pace. Four separate live Webinar question and answer sessions were scheduled in January and February to support this new training format.

Table 1.3: PSAE 2012 Test-Administration Schedule

	Test	Time (minutes)	Number of questions
Day 1	ACT English Test	45	75
	ACT Mathematics Test	60	60
	Break	*15*	—
	ACT Reading Test	35	40
	ACT Science Test	35	40
Day 2	ISBE-developed science	40	45
	WorkKeys *Applied Mathematics*	45	33
	Break	*15*	—
	WorkKeys *Reading for Information*	45	33

The Webinar consisted of three sections, each approximately one-half hour long. Part One provided an introduction to the PSAE as well as test administration policies and new information for 2012. Part Two included information for planning for the test days, maintaining the security of test materials, administering the test under standardized conditions, handling test irregularities, and providing accurate written information of test day procedures. Part Three included accommodations and additional Day 2 information.

When participants had completed their review of all three parts of the 2012 PSAE Training Webinar recording they could then attend a live Webinar question and answer session. The sessions covered the same material as the training sections so participants needed to only attend a single live session. In addition, the ACT *Supervisor's Manual for State Testing* and the *Day 2 Prairie State Achievement Examination Supervisor's Manual of Instructions* were posted on ISBE's website. These two manuals describe all procedures and requirements and include the verbal instructions that are read verbatim to students on test days. The manuals provide contact information so that testing staff can reach ACT and ISBE via telephone to consult about planning for the administration prior to the test days and to report testing irregularities on test days. On test days, ACT and ISBE staff were available by telephone beginning at 7:00 a.m. and 7:30 a.m, respectively.

Accommodations for Students with Disabilities

Appendix A contains detailed information and procedures for requesting accommodations on the PSAE.

ACT-Approved Accommodations

ACT provides test accommodations in accordance with Title III of the Americans with Disability Act (ADA). ACT's guiding principles for responding to requests from examinees for test accommodations:

- Requirements and procedures for test accommodations must ensure fairness for all candidates, both those seeking accommodations and those testing under standard conditions.
- Accommodations must be consistent with the Americans with Disabilities Act (ADA) requirements and appropriate and reasonable for the documented disability.
- Accommodations must not result in an undue burden, as that term is used under the ADA, or fundamentally alter that which the test is designed to measure.
- Documentation of the disability must meet guidelines that are considered to be appropriate by qualified professionals and must provide evidence that the disability substantially limits one or more major life activities. Applicants must also provide information about prior accommodations made in a similar setting, such as academic classes and test taking.

Review and Approval Process

Only examinees with professionally diagnosed and documented disabilities and who receive accommodations in school should apply for ACT-Approved Accommodations. Students who are receiving special education services described in a current individualized Education Program (IEP) or Section 504 Plan need to complete a Request for ACT-Approved Test Accommodations. Requests will be reviewed by ACT staff, and if appropriate, by other expert disability consultants, to ensure they meet ACT's established criteria and include the same supporting documentation required for approving all other ACT accommodations requests.

Examples of Accommodations

ACT-Approved Accommodations can include extended time, alternate test formats, stop-the-clock breaks, and authorization to test over multiple days. Examples of alternate test formats are audiocassettes or audio DVDs, Braille or large print.

ACT-Approved Accommodations are not available for students solely on the basis of limited English proficiency.

Reporting

ACT-Approved Accommodations that result in ACT scores are fully reportable to colleges, scholarship agencies, the NCAA and other entities in addition to being used for state testing purposes.

State-Allowed Accommodations

Students who do not meet the eligibility requirements for ACT-Approved Accommodations or those that were denied their request may apply for State-Allowed Accommodations.

Approval Process

Requests are made through ACT using the Application for State-Allowed Accommodations. ISBE has determined the process and guidelines for state-allowed accommodations for the PSAE.

Types of Accommodations

State-Allowed Accommodations include extended time, alternate test formats, stop-the-clock breaks, and authorization to test over multiple days. Examples of alternate test formats are audio cassettes or audio DVDs, Braille or large print. English language learners who do not have a disability but receive accommodations in school should request state allowed accommodations. Spanish video DVDs for Day 1 and Day 2 mathematics and science tests are available for eligible students. Additional information about this format can be found at **www.isbe.net/assessment/SpDVD.htm**. In addition, translated test instructions in 10 different languages are available for eligible students.

Reporting

Student ACT scores earned under state-allowed accommodations are NOT reportable to colleges, scholarship agencies, the NCAA and other entities; they can only be used for state purposes.

Key Difference Between ACT-Approved and State-Allowed Accommodations

Administrations of the ACT under ACT-Approved Accommodations result in scores that are fully reportable to colleges, scholarship agencies, and other entities in addition to being used for state testing purposes. Administrations of the ACT with State-Allowed Accommodations result in ACT scores appropriate for state use only.

Chapter 2
Validity Evidence for the Prairie State Achievement Examination

The Prairie State Achievement Examination (PSAE) measures student achievement relative to the Illinois Learning Standards. It measures the progress that schools have made in helping their students meet the Illinois Learning Standards, and it recognizes the excellent achievement of individual students whose scores qualify them for honors. The PSAE comprises three types of tests:

- A science test developed by Illinois teachers and curriculum experts working in cooperation with the Illinois State Board of Education (ISBE) and ACT,
- WorkKeys tests in reading and mathematics, and
- The ACT.

The PSAE and the Illinois Learning Standards

The PSAE is required by Illinois law to measure student performance in three academic areas: reading, mathematics, and science. In addition to meeting the state requirements, the PSAE must fulfill the requirements of the federal Elementary and Secondary Education Act, which requires states to develop and adopt (1) challenging content and student performance standards and (2) a set of high-quality student assessments to be used to determine the yearly performance of each public school.

With passage of the current PSAE legislation in 1999, ISBE staff were directed to explore the possibility of developing an examination to fulfill state and federal testing requirements for high school students that comprised three types of assessments: a college-placement assessment; assessments used for job placement; and ISBE-developed assessments to cover the Illinois Learning Standards not sufficiently covered by the other assessments.

For the proposed PSAE to meet both the state and federal requirements, it had to assess the three required academic areas and be aligned with the Illinois Learning Standards. No single assessment can effectively measure every one of the Standards. Table 2.1 summarizes the Illinois Learning Standards measured by the PSAE. The match to the Illinois Learning Standards was the foremost consideration for selecting components of the PSAE. To determine how well the ACT, two WorkKeys assessments, and the ISBE-developed science test covered the necessary content, ISBE conducted reviews that compared the contents of these tests with the Illinois Learning Standards.

Prior to the first PSAE administration in 2001, ISBE reviewed the ACT and a study that ACT had previously done that compared the ACT to the Illinois Learning Standards. ISBE also reviewed two WorkKeys assessments in light of the Illinois Learning Standards. The results of these reviews showed that the ACT coupled with the ISBE-developed science test and the WorkKeys reading and mathematics assessments provided a good match to the Illinois Learning Standards. ISBE staff also commissioned independent reviews to verify that a PSAE composed of the ACT, two WorkKeys assessments, and the ISBE-developed science test match the Illinois Learning Standards that it is intended to measure. The studies that reviewed each component of the PSAE to the Illinois Learning Standards are discussed in the following sections.

The ACT's Match to the Illinois Learning Standards

The ACT is a curriculum-based assessment program. Test specifications for each of the tests that make up the ACT are based on studies done every three to four years by ACT of curricula in use throughout the United States. The ACT curricula studies consist of reviewing the state educational standards of the 49 states that have established such standards; consulting with college and high school teachers and administrators, subject-area experts, and curriculum specialists; monitoring published commentaries on education in the United States; reviewing widely used high school and college textbooks; and surveying practicing educators about classroom methods and instructional emphases. Using these data, ACT identifies the knowledge and skills students need to learn in high school to be prepared for college. See ACT 2009 for the results of the most recent ACT National Curriculum Survey. The foundation of the ACT is in the curriculum; thus, since state standards are intended to

Table 2.1: How the PSAE Measures Student Progress Toward Meeting the Illinois Learning Standards (ILS)

PSAE tests	What the ILS require	How the PSAE measures the ILS
Reading	Ability to read with fluency and understanding and to comprehend a broad range of reading materials (ILS 1A–C).	Provides comprehensive assessment of reading skills: • Academic reading passages that include prose fiction, humanities, social science, and natural science • Work-related informational pieces, such as policies, bulletins, letters, manuals, and governmental regulations • Multiple-choice questions that require students to reference the text and think critically
Mathematics	Understanding and ability to apply knowledge of number sense, estimation, and arithmetic (ILS 6A–D; 7A, B; 8C); algebra (8A–D); geometry and trigonometry (9A–D); measurement (7C); and data organization and probability (10A–C).	Provides comprehensive assessment of mathematics knowledge and skills: • Assesses mathematical skills acquired in courses taken through grade 11 • Academic and work-related content assessed through increasingly complex tasks • Multiple-choice questions require mathematical reasoning to solve practical problems • Approved calculators may be used, and complex formulas are provided
Science	Understanding and ability to apply knowledge of experimental design (ILS 11A) and technological design (11B), including how to conduct controlled experiments and analyze and present the results; life sciences (12A, B), chemistry (12C), physics (12D), Earth science (12E), and space science (12F); laboratory safety, valid sources of data, and ethical research practices (13A); and historical interactions between science, technology, and society (13B).	Measures scientific knowledge and its application: • Interpretation, analysis, evaluation, reasoning, and problem-solving skills • Science inquiry; life, physical, and Earth and space sciences; and science, technology, and society • Multiple-choice questions that assess the ability of students to use critical thinking skills to evaluate information provided on the test

define what teachers should be teaching, the ACT has a relationship to state standards.

In addition, ACT staff have completed matches between the ACT and the standards of more than 40 states, including the Illinois Learning Standards. ISBE reviewed ACT's study comparing the skills assessed on the ACT with the Standards. The first ACT study was conducted in two parts: Part 1, conducted in 1999, looked at the Illinois Learning Standards to determine which of them were measured by the ACT. The results of this study showed that in language arts (State Goals 1, 2, and 3), five of the six Illinois Learning Standards under reading and writing are covered on the ACT. In mathematics (State Goals 6, 7, 8, 9, and 10), 16 of the 18 Illinois Learning Standards are covered by the ACT. In science, State Goal 11 matches well with the knowledge and skills measured by the ACT Science Test. Part 2 of the study, conducted in 2000, looked at the ACT College Readiness Standards® (the knowledge and skills students in various score ranges of the ACT are likely to have attained) to determine if what is measured by the ACT is part of the Illinois Learning Standards. The results of Part 2 of this study showed that nearly all of the ACT College Readiness Standards (formerly known as ACT's Standards for Transition) are subsumed under the Illinois Learning Standards. The detailed results of both parts of the ACT study are summarized in two reports: *Comparison of the Illinois Learning Standards to the ACT Assessment, PLAN, and EXPLORE* (ACT, 1999) and *Comparison of the Illinois Learning Standards to the ACT Assessment Standards for Transition* (ACT, 2000). In 2006, ACT staff again examined the match between the Illinois Learning Standards and the ACT, PLAN, and EXPLORE and found similar results to the previous study (ACT, 2006).

To conduct its own review of the relationship of the Illinois Learning Standards to the ACT, ISBE convened meetings of Illinois educators who were engaged in instruction aligned with the Illinois Learning Standards to review the match between the ACT and the Illinois Learning Standards. The results of this review also showed that there is substantial agreement between the

ACT and the Illinois Learning Standards. The reviews conducted by the Illinois educators in February 2000 are discussed in detail on pages 7–8 of this manual.

The WorkKeys Match to the Illinois Learning Standards

The WorkKeys *Reading for Information* and *Applied Mathematics* assessments were selected because of their match to the "Applications of Learning" sections of the Illinois Learning Standards; that is, the WorkKeys assessments provide a measure of whether students can apply classroom knowledge and skills to situations necessary for employment and successful living in the twenty-first century.

The WorkKeys assessments used in the PSAE serve two purposes:

1. The two assessments increase the range of acquired abilities assessed by the PSAE, and
2. Students can use these assessments to identify the workplace skills they possess and the skills they need to acquire.

Several comparisons of the WorkKeys skill descriptions and the Illinois Learning Standards have been conducted. In February 2000, a match analysis was conducted by ACT staff and reviewed by ISBE staff. The WorkKeys *Reading for Information* assessment was found to match all the components of Illinois State Goal 1. The WorkKeys *Applied Mathematics* assessment was found to match components in Illinois State Goals 6, 7, 8, 9, and 10. Also in February 2000, ISBE convened meetings of Illinois educators who were engaged in instruction based on the Illinois Learning Standards to review the match between the WorkKeys assessments and the Illinois Learning Standards. The results of the review by Illinois educators also showed that there is significant agreement between the WorkKeys *Applied Mathematics* and *Reading for Information* assessments and the Illinois Learning Standards. The reviews conducted by the Illinois educators are discussed in the following section.

Review of PSAE Alignment to the Illinois Learning Standards by Illinois Educators

Three meetings were held in late February 2000 to conduct reviews of the alignment of the ACT Test, the WorkKeys assessments, and the ISBE-developed tests (which at the time included a science test and a writing test) to the Illinois Learning Standards. The language arts meeting was held in Springfield on February 25, 2000, with 25 high school language arts teachers. The mathematics meeting was held in Champaign on February 26, 2000, with 25 high school mathematics teachers. The science meeting was held in Springfield on February 29, 2000, with 15 high school science teachers. All participating teachers had previously served on ISBE assessment advisory committees or participated in the development and review of previous ISBE-developed assessments. Each of the three meetings started at 8:30 a.m. and lasted until approximately 3:30 p.m.

At each of the three meetings the teachers first listened to presentations from ISBE Assessment Division Administrator, Dr. Carmen Chapman Pfeiffer, and from ACT representatives who were content specialists for the subject under review. Teachers were given copies of a released ACT Test, the WorkKeys assessment relevant to their subject, and the ISBE-developed pilot test relevant to their subject. They also received the results of the ACT review of the ACT Test's alignment with the Illinois Learning Standards and worksheets that listed each Standard with space in which they could indicate how well each of the three assessments covered each Standard.

After the group presentations, the teachers formed small discussion groups. They reviewed the test materials in light of the Illinois Learning Standards for their subject, engaged in discussions, and then completed a form that summarized the coverage of the Illinois Learning Standards by the ACT Test and WorkKeys components and the ISBE-developed test.

Results of the Language Arts Review by Illinois Educators

The Illinois English teachers found that the ACT English Test thoroughly covers conventions (punctuation, grammar and usage, and sentence structure) and editing skills (strategy, organization, and style). The English teachers found there to be a good match between the ACT Reading Test and the Illinois Learning Standards for English that specifically address reading.

The "real-world documents" in WorkKeys *Reading for Information* are used to assess communication skills needed in the workplace. This connection to the work-place addresses the "Applications of Learning" that are part of the Illinois Learning Standards for each subject.

Results of the Mathematics Review by Illinois Educators

The mathematics teachers found there to be a good match between the ACT Mathematics Test and the Illinois Learning Standards for mathematics. The ACT Mathematic Test subscore areas are similar to the standard-set groupings that ISBE staff generated for mathematics.

The “real-world documents” in WorkKeys *Applied Mathematics* are used to assess skill in using mathematical reasoning to solve work-related problems. This connection to the workplace addresses the Application of Learning for mathematics, which states, “…particularly in an occupational setting, the [mathematics] problems are non-routine and require some imagination and careful reasoning to solve. Students must have experience with a wide variety of problem-solving methods and opportunities for solving a wide range of problems.”

Results of the Science Review by Illinois Educators

The science educators found that the ACT Science Test aligns well with ILS 11A, scientific inquiry, and shows application to the content areas covered by Illinois Learning Standards in Goal 12, which include life sciences, chemistry, physics, and Earth and space science. While the ACT Science Test has applications to Goal 12 Standards, the teachers concluded that it does not require students to demonstrate sufficient specific understanding of the content areas. Other Illinois Learning Standards not specifically covered are ILS 11B, technological design; ILS 13A, the accepted practice of science; and ILS 13B, science and technology in society. The ISBE-developed science test covers the Standards not included as part of the ACT Science Test.

Independent Reviews of the PSAE Assessments

In 2000, ISBE contracted with reading and mathematics experts for review of the PSAE reading and mathematics tests and their alignment with the Illinois Learning Standards. Donna Ogle and Kenneth Hunter reviewed the reading tests; John A. Dossey and Sharon Soucy McCrone reviewed the mathematics tests. Detailed results of these reviews can be found in Appendix B.

As part of its ongoing efforts to evaluate the alignment of the Illinois Learning Standards with the PSAE, in February 2006, ISBE also commissioned Norman Webb to conduct an independent alignment study of the PSAE Reading, Mathematics, and Science components to the Illinois Learning Standards (see Webb 2006a, 2006b, and 2006c).

Reviews conducted to date of the alignment between the PSAE components and the Illinois Learning Standards support ISBE’s conclusion that although a few weaknesses exist, overall the PSAE adequately covers the Illinois Learning Standards in reading, writing, mathematics, and science.

Additional Validity Evidence

ACT and WorkKeys as Part of the PSAE

The ACT was developed as a college entrance examination; consequently, educators and others have questioned its appropriateness for all high school students, not all of whom will attend college. This section addresses the following questions: Is the ACT an appropriate assessment for all high school students? Are the WorkKeys assessments appropriate for all students in high school, even those planning to attend college immediately after high school?

To provide evidence for the content validity of the ACT and WorkKeys assessments as part of the Illinois statewide assessment program—specifically as a possible component of the PSAE—ISBE and ACT engaged in a rigorous evaluation process guided by ACT’s eight necessary conditions.

Condition 1: *The ACT and WorkKeys assessments must measure the state’s standards.* The PSAE was established to measure the Illinois Learning Standards, so a necessary precondition to use of the ACT and WorkKeys assessments as part of the PSAE was to ensure that the knowledge and skills measured by the ACT and WorkKeys assessments are included in the Illinois Learning Standards. Several different evaluation studies were conducted, one by ACT and several by ISBE. These are described in this chapter of this manual.

Condition 2: *The use of the ACT and WorkKeys assessments should be consistent with the intended outcomes of the statewide assessment program.* The PSAE was established to show the progress that schools, districts, and the state have made toward meeting the Illinois Learning Standards in four subjects: reading, mathematics, science, and writing. The PSAE also measures each student’s academic achievement with respect to the Illinois Learning Standards and provides an opportunity for individuals to receive recognition for excellent performance in one or more of these subjects.

The Illinois Learning Standards are statements of the specific knowledge and skills that every public school student should learn in school. The Illinois Standards Project began in 1995 and was completed in 1997. Thousands of Illinois citizens—teachers, parents, school administrators, employers, community leaders, and representatives of higher education—identified what they believe students will need to know and be able to do when they graduate from high school. The Illinois Learning Standards were developed to be essential to both entry-level jobs and post–high school education. Whether students intend to go directly to work or plan to attend a vocational or technical school, junior college, or four-year college, those who meet the Illinois Learning Standards will have the academic background they need to compete successfully.

Because ISBE wanted the PSAE to have value for individual students, the program was designed to include three types of measures: the ACT Test, which can also be used for college admissions; two WorkKeys tests that measure skills in mathematics and reading that employers believe are critical for job success and can be included in a student's work portfolio; and an ISBE-developed test in science to ensure comprehensive coverage of the Illinois Learning Standards.

The ACT measures academic strengths and weaknesses relative to college readiness. Students considering college right after high school may use their ACT scores for college admissions. Others who decide to return to school after they have worked for a time can also use their scores for admissions. High school students may use their WorkKeys scores to identify the reading and mathematics skills they have developed and those they need to acquire to qualify for various jobs. The ISBE-developed science test covers skills and knowledge that are not specifically addressed by the ACT Test and WorkKeys assessments but that are necessary for students to be successful in their roles as citizens and participants in our society.

The goals of the PSAE and the purposes of the ACT Test and WorkKeys are philosophically consistent: both programs are committed to providing students with information that has value independent of the state's use of the results for school accountability.

Condition 3: *Neither the ACT nor WorkKeys assessments should be used by themselves as the sole criterion in making high-stakes decisions about students.* From the outset, it was clear that the results of the PSAE would not be used as a high school graduation requirement. Section 2-3.64 of the Illinois School Code states, "A student who successfully completes all other applicable high school graduation requirements but fails to receive a score on the Prairie State Achievement Examination that qualifies the student for receipt of a Prairie State Achievement Award shall nevertheless qualify for the receipt of a regular high school diploma" (105 ILCS 5/2-3.64). Rather, the results are being used by high school teachers, curriculum coordinators, and administrators to evaluate the effectiveness of their curricula and instruction in helping students acquire the knowledge and skills defined by the Illinois Learning Standards. Students who earn qualifying scores in one or more of the PSAE subjects receive a Prairie State Achievement Award, but that award is not used to make any high-stakes decisions about students.

Condition 4: *Neither the ACT Test nor WorkKeys assessments should be used as the sole criterion in making high-stakes decisions about school or teacher effectiveness.* Consistent with the purposes of the PSAE, the information provided through the program is used to evaluate the progress schools and districts have made in meeting the Illinois Learning Standards. ISBE also is using this information to help identify paths for improvement for those schools not making adequate yearly progress. Neither the ACT scores nor WorkKeys scores are used as the sole criterion in these evaluations.

Condition 5: *Opportunities must be provided to inform students and parents about what the ACT Test and WorkKeys assessments measure, what the scores mean, and how the scores can help students prepare for what they want to do after high school.* Orientation workshops were initially conducted throughout the state on September 18–28, 2000, to fully brief high school educators on the new program and how to use the results. To summarize the information provided in the workshops, each high school receives a supply of the *PSAE Teacher's Handbook*, which contains the test administration schedule, test preparation information, and a comprehensive description and review of all the PSAE tests, including sample questions.

In the first year of the program, ISBE purchased ACT and WorkKeys materials, including *ACTive Prep: The Official Electronic Guide to the ACT Assessment®, ACT College Readiness Standards, ACT Test Preparation Reference Manual, Getting into the ACT, WorkKeys Occupational Profiles, WorkKeys Targets for Instruction: Reading for Information,* and *WorkKeys Targets for Instruction: Applied Mathematics.* These materials were shipped to each high school in September 2000. Other materials were provided free of charge, including *Preparing for the ACT Assessment* and *Preparing for the Work Keys Assessments*. Every year, high schools also receive information pertaining to the PSAE as a whole and the ISBE-developed science test, including the *PSAE Parent Brochure*, the *PSAE Day 2 Overview and Preparation Guide*, and the *PSAE Teacher's Handbook.* All of these materials help familiarize teachers, students, and parents with the component tests, test content, and test format.

ISBE and ACT believe that the ACT Test and WorkKeys assessments provide information that can help all students. For example, students who are considering going to college after high school can use their scores on the ACT Test to evaluate their readiness for college. Scores obtained on the ACT taken as part of the PSAE can be submitted to colleges throughout the United States for admission and course placement just as can scores obtained on a national ACT test date. Also, students who are not considering college may decide to do so after taking the ACT and receiving their scores. Students who plan to work or go into technical or other training after high school may use the ACT scores and WorkKeys assessments scores as feedback about their relative strengths and weaknesses so that they can be prepared to achieve their goals. Because the ACT and WorkKeys assessments measure achievement in critical areas needed throughout life, the scores offer valuable information that can be used in positive ways regardless of students' future plans.

The ACT provides both normative interpretations of scores (interpretations of performance relative to the performance of other students) and standards-based interpretations of scores (interpretations of performance described in terms of content and skill standards) through the ACT College Readiness Standards. Some students may want to compare their performance to the performance of others having similar postsecondary plans; others may prefer to examine their performance relative to what they know and can do and what they need to learn to achieve their postsecondary goals. WorkKeys assessments are criterion-referenced, so score reports differ somewhat. However, students can use report information, score interpretation guides, Job Skills comparison charts, and Occupational Profiles to guide their important life decisions. Thus, all students can use the ACT Test and WorkKeys information to prepare themselves, no matter what they decide to do after high school.

Condition 6: *A statewide assessment program will be effective only when teachers and administrators have opportunities to learn more about the assessments, what they measure, how they are developed, and how the results relate to instruction.* This applies to the PSAE as a whole and to the ACT Test and WorkKeys assessments that are included in the PSAE. All of the steps described under Condition 5 were also intended to help teachers and administrators understand the PSAE program and to make informed uses of the results. This information, as well as other information about score interpretation and use, was the focus of combined ISBE-ACT workshops for curriculum coordinators held in September 2001 and workshops for guidance counselors and administrators held in November 2001.

Condition 7: *The ACT Test and WorkKeys assessments must be administered under secure, standardized conditions that will provide each student a fair and equitable opportunity to demonstrate what he or she has learned and assure the integrity of the test scores to those who interpret and use the results.* It is critically important that the PSAE, including the ACT Test and WorkKeys assessments, be administered under secure, standardized conditions. To ensure proper implementation of the standard testing requirements for the PSAE, educators designated as test supervisors, back-up test supervisors, or test accommodations coordinators at their schools were trained as described in this manual.

ISBE and ACT staff conduct several in-person site audits on the test day to observe the administration. A review of these audit reports and other test day documentation submitted from the test sites indicate that the overall test experience was very similar to that of a national ACT test day. In the few cases of reported timing shortages or severe distractions, students were given the option of testing on the scheduled makeup date two weeks later.

Condition 8: *When the ACT Test and WorkKeys scores are combined with other statewide assessment measures, it is important that students derive maximum value from them—both as one of several measures of their achievement related to statewide goals and as an independent indicator of their college and workplace readiness.*

The PSAE was designed to provide scores that reflect the combined PSAE measures as well as a standard ACT student report. If the ACT Test is used as one of several measures of student achievement included in the PSAE, the ACT scores may be combined with the scores of other measures to form PSAE scores reflecting overall student performance in the subject areas measured. These scores have meaning and value within the statewide assessment context and should inform both instruction and individual improvement within the classroom setting. Likewise, the WorkKeys scores provide valuable information related to training needs. Beyond their use as one of several measures within the PSAE, ACT scores also have independent value to students when reported to the schools and colleges requested by students. The ACT scores can be used by students for admission to college or as an early indication of the areas in which students may want to take additional course work before applying to college.

Because ACT scores are reported both independently to schools and colleges and as part of the PSAE, Illinois students are more likely to receive the full and complete benefits of each. The PSAE score report includes three PSAE scores, one for each of the three PSAE subjects: reading, mathematics, science, and writing. The ACT student report contains scores for each of the four ACT tests, eight subscores, and a composite score. ACT scores must not be included on student transcripts without the permission of the student or of the student's parent or guardian if the student is not 18 years of age. The WorkKeys score reports contain scores for both *Reading for Information* and *Applied Mathematics* skills as well as suggestions for improvement. They may be used at the student's discretion for workplace and training applications.

Colleges and universities throughout the United States, including the Ivy League schools, have indicated their willingness to use ACT scores reported from state testing. In addition, the Illinois Board of Higher Education, the Illinois Community College Association, and the Illinois Student Assistance Commission (ISAC) have fully endorsed and used ACT scores deriving from PSAE testing. Employers accept WorkKeys scores from PSAE testing as well.

Criterion-Related Validity Evidence for PSAE Science

These analyses examined the criterion-related validity of PSAE science scale scores. Using data from the 2008 spring PSAE administration, three external criterion variables related to high school course work were selected: 1) science course grades, 2) number of semesters students have taken science courses, and 3) whether students have taken advanced science courses. These three variables were based on self-reported student information.

Average PSAE science scale scores, grouped by each of the criterion variables, are presented in Tables 2.2, 2.3, and 2.4, respectively. As shown, the average PSAE science score increases as the course grade increases for the subjects of general science, biology, chemistry, and physics. Students tend to have higher PSAE scores if they have taken science courses for a longer period of time, and students who have taken advanced science courses score higher than students who have not. The criterion-related validity of PSAE science is supported by this evidence, which shows a positive relationship between students' scientific knowledge and skills and their performance on the PSAE science test.

Table 2.2: Average PSAE Science Scale Scores, by Science Course Grades

General Science course grade	PSAE	Biology course grade	PSAE	Chemistry course grade	PSAE	Physics course grade	PSAE
F	143	F	146	F	151	F	152
D	145	D	149	D	153	D	153
C	149	C	153	C	158	C	158
B	155	B	160	B	165	B	167
A	164	A	168	A	171	A	174

Table 2.3: Average PSAE Science Scale Scores, by Semesters of Science

Number of semesters of science	Mean PSAE science score
1	140
2	143
3	146
4	149
5	150
6	158
7	157
8	167

Table 2.4: Average PSAE Science Scale Scores, by Students with Advanced Courses in Natural Sciences

AP, accelerated, or honors courses in natural sciences	Mean PSAE science score
Yes	168
No	155

Descriptions of the Components of the PSAE

To fully measure the Illinois Learning Standards, the PSAE is comprised of multiple assessments, as presented in Chapter 1. The three types of tests making up the components are the ISBE-developed science test, two WorkKeys assessments, and the ACT. Each type of test is further described below in terms of what each test measures, how each test is developed, and the technical characteristics of each test.

The ISBE-Developed Science Test

The PSAE includes an ISBE-developed assessment in science. The ISBE-developed science test is designed to assess the Illinois Learning Standards validly and fairly.

Description of the ISBE-Developed Science Test

The selection of items and assembly of each test is guided by a set of test specifications. These specifications were developed by Illinois educators to help ensure that test content is aligned to the purposes, objectives, and skills framed by the Illinois Learning Standards.

Illinois teachers and administrators participate in all phases of the test development process: item writing, item editing, and item data review. ISBE convenes a series of advisory committees to ensure that test development is continually informed and guided by the recommendations of content authorities, measurement specialists, and practitioners. The following evaluation criteria are applied to all assessment material used in the ISBE-developed science test:

Content. Every item is screened for alignment with the Illinois Learning Standards, grade-level appropriateness, importance, and clarity. Incorrect choices (for multiple-choice items) are reviewed for plausibility. The complexity of the text of the questions is kept to the minimum necessary to state the problem.

Difficulty. Items are pilot tested on large samples of students to develop a statistical profile for each item before their inclusion in the PSAE. Items that are too easy or too difficult and, therefore, provide little or no information are omitted.

Discrimination. Point-biserial (i.e., item-test) correlations evaluate the extent to which an item distinguishes between less proficient and more proficient students. Test items with the highest point-biserial values are selected to use on test forms, with a minimum acceptable value of 0.20.

Fairness. Test items and forms undergo regular sensitivity reviews and statistical analyses to ensure that all materials meet fairness criteria with respect to the cultural and ethnic diversity of Illinois public schools.

The ISBE-developed component of the PSAE science assessment consists of 40 single-right-answer, multiple-choice items. The score from the ISBE-developed science test items are combined with the scores from the ACT Science Test to produce the PSAE science score. In addition to the overall PSAE science score, results are reported for the ISBE-developed science test and for the ACT Science Test. The ISBE-developed science test scale was defined by letting 70 represent the average proficiency of the first-year test population. Every unit on the scale represents 1/10 of the standard deviation of proficiency scores for the first-year population. In other words, the first-year mean and standard deviation of scale scores are 70 and 10, respectively.

The Productive Thinking Scale (PTS) is used to evaluate the quality of items used in the ISBE-developed component of the PSAE science assessment. It is hierarchical with respect to the production of knowledge and independent of an item's difficulty. Four cognitive skills define the hierarchy of productive thinking in generating scientific knowledge. Each skill applies to both content (knowledge) and process (research methods):

1. recall of conventions, whether names or norms;
2. reproduction of empirical facts or methodological tools and steps;

3. production of solutions to problems or research designs; and
4. creation of new theories and methods.

The PTS further subdivides reproduction and production into secondary processes, for a total of six levels of productive thinking on a scale from low level (recall of conventional uses) to high level (creation of new theory).

> ### *Illinois State Goals in Science*
>
> Illinois State Goals 11, 12, and 13 address science. The Illinois Learning Standards (ILS) within these goals inform one another and depend upon one another for meaning. The ISBE-developed component of the PSAE science assessment is designed to measure the following Illinois Learning Standards.
>
> ***State Goal 11:*** *Understand the process of scientific inquiry and technological design to investigate questions, conduct experiments and solve problems.*
>
> **ILS 11A.** Know and apply the concepts, principles and processes of scientific inquiry.
> **ILS 11B.** Know and apply the concepts, principles and processes of technological design.
>
> ***State Goal 12:*** *Understand the fundamental concepts, principles and interconnections of the life, physical and earth/space sciences.*
>
> **ILS 12A.** Know and apply concepts that explain how living things function, adapt and change.
> **ILS 12B.** Know and apply concepts that describe how living things interact with each other and with their environment.
> **ILS 12C.** Know and apply concepts that describe properties of matter and energy and the interactions between them.
> **ILS 12D.** Know and apply concepts that describe force and motion and the principles that explain them.
> **ILS 12E.** Know and apply concepts that describe the features and processes of the earth and its resources.
> **ILS 12F.** Know and apply concepts that explain the composition and structure of the universe and Earth's place in it.
>
> ***State Goal 13:*** *Understand the relationships among science, technology, and society in historical and contemporary contexts.*
>
> **ILS 13A.** Know and apply the accepted practices of science.
> **ILS 13B.** Know and apply concepts that describe the interaction between science, technology, and society.

Based on estimates of the thought processes that most students must use to answer an item, each item is ranked with respect to the level of cognitive skill it requires. Items are also examined to determine whether there is a distribution within tests of items across the standards: earth science, physical science, and life science.

Reliability of the ISBE-Developed Science Test

Test reliability indicates the extent to which differences in test scores reflect real differences in the ability being measured and, thus, the consistency of test scores across some change of condition, such as a change of test items or a change of time. Different reliability coefficients result from different changes in testing conditions.

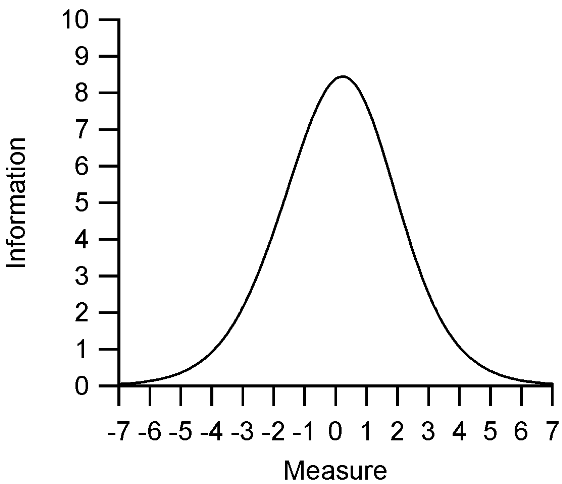
The reliability of the ISBE-developed science test is estimated by coefficient alpha. Coefficient alpha is an internal consistency reliability coefficient because it can be calculated from one administration of the test and depends on the inter-relatedness of the items. It is the average item inter-relatedness, and it reflects how consistently the items measure the tested construct. The value of coefficient alpha for the 2012 ISBE-developed science test was 0.82 based on a sample size of 127,440.

The value is derived from the total test population. For well-constructed achievement tests, internal consistency reliability coefficients typically exceed 0.90. Internal consistency estimates are influenced both by the interrelatedness of test items and the number of test items. Since the 40-item ISBE-developed science test represents only half the PSAE science assessment, internal consistency is slightly lower than is typical for ISAT science tests.

The reliability coefficient reported is derived within the context of classical test theory (CTT) and provides a single measure of precision for the entire test. Within the context of item response theory (IRT), it is possible to measure the relative precision of the test at different points on the scale. Figure 2.1 presents the test information functions for the ISBE-developed science test. Note that the test information function is computed from the test as a whole, although ISBE-developed science test scale scores are calculated by averaging four subscale scores.

A second way of evaluating precision from the IRT perspective is in terms of how well the test as a whole separates persons. The ratio of the standard deviation of ability estimates, after subtracting from their observed variance the error variance attributable to their standard errors of measurement, to the root mean square standard error computed over persons provides this index (Wright & Stone, 1979). The person separation value for the 2012 ISBE-developed science test is 2.13. Values around 3.00 and above are desirable for achievement tests such as the ISBE-developed component of the PSAE assessment. Because the ISBE-developed science test comprises only 40 items and represents only half the PSAE science assessment score, the person separation estimate was not expected to be at an optimal level.

Figure 2.1: 2012 ISBE-Developed Science Test Information Function



Scaling Procedures for the ISBE-Developed Science Test

Overall PSAE scores are reported on a standard score scale on which individual student scores range between 120 and 200, regardless of the characteristics of the raw score distribution. Each scale is defined by letting 160 represent the average proficiency and 15 the standard deviation of a sample of 10,554 students from the total first-year test population. The scaling analyses for these tests were conducted on this sample.

The statistical fit of the one-parameter logistic (1PL) or Rasch model to the ISBE-developed science and social science tests has been examined previously and found to be satisfactory. The 1PL model uses only the item difficulty and the person's proficiency level to describe the probability of a correct response to an item. The 1PL model is the simplest of currently available IRT models and is perhaps the one in widest use today.

Table 2.5 shows results of the Rasch calibrations for the science test. Column 1 shows the item number within the test booklet. Column 2 shows the Rasch difficulties and column 3 shows the standard error of the difficulty estimate (S_{ed}). The next two columns present statistics designed to assess how well the test fits the IRT model. Both are standardized, mean-square statistics with an expected value of 1.00 (indicating perfect fit). The first, "Infit," is more sensitive to departures from model fit when item difficulty and person ability are close. The second, "Outfit," is more sensitive to model fit when item difficulty and person ability are far apart. The last column shows the point-biserial correlation between the item and the rest of the items in the test.

Table 2.5: Results of the 2001 Rasch Calibration Process for Science

Item	Difficulty	S_{ed}	Infit	Outfit	rpb
1	0.36	0.02	0.94	0.91	0.46
2	–0.42	0.02	1.14	1.22	0.22
3	–0.66	0.03	1.06	1.11	0.28
4	2.71	0.03	1.18	1.89	0.12
5	–0.82	0.03	0.96	0.97	0.36
6	1.31	0.02	1.02	1.05	0.39
7	0.13	0.02	1.00	0.99	0.39
8	–1.33	0.03	0.92	0.82	0.37
9	–0.51	0.02	1.09	1.18	0.26
10	0.21	0.02	1.03	1.04	0.37
11	–0.80	0.03	1.01	0.97	0.33
12	0.70	0.02	0.93	0.92	0.47
13	–0.50	0.02	1.02	1.12	0.32
14	0.96	0.02	1.08	1.11	0.34
15	0.22	0.02	1.04	1.06	0.35
16	1.13	0.02	0.90	0.89	0.50
17	0.18	0.02	0.93	0.88	0.46
18	–0.42	0.02	0.92	0.83	0.44
19	0.88	0.02	1.08	1.11	0.34
20	1.17	0.02	0.92	0.91	0.48
21	1.58	0.02	1.07	1.16	0.33
22	1.00	0.02	1.09	1.14	0.32
23	–0.33	0.02	1.02	1.07	0.34
24	–1.36	0.03	0.90	0.70	0.40
25	–0.12	0.02	1.02	1.04	0.35
26	0.07	0.02	1.02	1.00	0.37
27	0.46	0.02	1.00	0.98	0.41
28	–1.08	0.03	0.91	0.81	0.39
29	0.27	0.02	0.98	0.97	0.41
30	0.43	0.02	0.99	0.97	0.41
31	0.38	0.02	0.99	0.98	0.41
32	–0.74	0.03	0.98	1.09	0.34
33	–2.23	0.04	0.90	0.61	0.33
34	0.14	0.02	1.14	1.26	0.25
35	–0.52	0.02	0.98	0.99	0.37
36	–0.78	0.03	0.95	0.97	0.37
37	–1.39	0.03	0.98	1.14	0.28
38	–0.83	0.03	0.87	0.74	0.46
39	0.20	0.02	0.91	0.87	0.48
40	0.37	0.02	0.92	0.89	0.47

After calibration, the ISBE-developed science component was scaled to a mean of 70 and a standard deviation of 10 within the total test population. The scaling constants used to transform the Rasch proficiency estimates to the reporting scales are shown in Table 2.6.

Table 2.6: PSAE Scaling Constants

	Slope	Intercept
ISBE-Developed Science	9.4628	63.8827

The WorkKeys Assessments Components: *Reading for Information* and *Applied Mathematics*

In recent years, members of the business community as well as the general public have indicated concern that American workers, both current and future, lack the workplace skills needed to meet the challenges of rapidly evolving technical advances, organizational restructuring, and global economic competition. New jobs often require workers coming from high schools or postsecondary programs to have strong problem-solving and communication skills. Current trends in basic skill deficiencies indicate that American businesses will soon be spending more than $25 billion a year on remedial training programs for new employees.

ACT designed WorkKeys to solve this problem. The system serves businesses, workers, educators, and learners. As part of the development process, ACT listened to employers, educators, and experts in employment and training requirements to find out which employability skills are crucial in most jobs. Based on their insights, ACT developed the following WorkKeys skill areas: *Applied Technology*, *Applied Mathematics*, *Business Writing*, *Listening*, *Locating Information*, *Reading for Information*, *Teamwork*, *Workplace Observation*, and *Writing*. Personal skills assessments have also recently been developed in the areas of *Performance*, *Talent*, and *Fit*.

Each skill area has its own skill scale that measures both the skill requirements of specified jobs and the employability skills of individuals. Before WorkKeys, scales could not easily measure both the skills a person has and the skills a job needs. Each WorkKeys skill scale describes a set of skill levels. This makes it possible to determine the proficiency levels students and workers already have and to design job-training programs that can help them meet the demands of the jobs they want. The WorkKeys system is based on the assumption that people who want to improve their skills can do so if they have enough time and appropriate instruction. Showing a direct connection between job requirements and education and training has a positive effect on learner persistence and achievement.

The WorkKeys Assessment Development Process

WorkKeys assessments are designed to cover a range of skills that is not too narrow and not too wide. If too narrow, a huge battery of tests would be needed to measure skills accurately; and if too wide, the number of items needed for validation would make the assessment too long and time-consuming. Thus, the WorkKeys assessments are designed to meet the following criteria:

- The way a skill is assessed is generally congruent with the way the skill is used in the workplace.
- The lowest level assessed is at approximately the lowest level for which an employer would be interested in setting a standard.
- The highest level assessed is at approximately the level beyond which specialized training would be required.
- The steps between the lowest and highest levels are large enough to be distinguished and small enough to have practical value in documenting workplace skills.
- The assessments are sufficiently reliable for high-stakes decision making.
- The assessments can be validated against empirical criteria.
- The assessments are feasible with respect to cost, administration time, and complexity.

The development process for a WorkKeys assessment consists of five phases: skill definition, test specifications development, prototyping, pretesting, and construction of operational forms. The process used to develop the WorkKeys multiple-choice test items is similar to that used for many standardized assessments including others developed by ACT (Anastasi, 1982; Crocker & Algina, 1986). Both stimuli and response alternatives meet basic requirements associated with high-quality skills.

Skill Definition

Before constructing the WorkKeys assessments, ACT defines the content domains and develops hierarchical WorkKeys skill descriptions. This process typically begins with a panel made up of employers, educators, and

ACT staff. The panel first develops a broad definition of a skill area and identifies the lowest and highest level of the skill that is worthwhile to measure. The panel then identifies examples of tasks within this broadly defined skill domain and narrows that domain to those examples that are important for job performance across a wide range of jobs. Next, the tasks are organized into "strands," which are aspects of the general skill domain, or skill area that pertain to a singular concept to be measured. The strands assessed in *Reading for Information*, for example, include "choosing main ideas or details," "understanding word meanings," "applying instructions," and "applying information and reasoning."

The strands are also divided into levels based on the variables believed to cause a task to be more or less difficult. In general, at the low end of a strand a few simple things must be attended to, whereas at the high end, many things must be attended to and a person must process information to apply it to more complex situations. In the "applying instructions" strand of *Reading for Information*, for example, employees need only apply instructions to clearly described situations at the lower levels. At the higher levels, however, employees must not only understand instructions in which the wording is more complex, meanings are more subtle, and multiple steps and conditionals are involved, but must also apply these instructions to new situations.

Test Specifications

Using the skill definitions described above, the ACT WorkKeys development team works on the specifications, outlining in more detail the skills the assessment will measure and how the items will become more complex as the skill levels increase. Each level is defined in terms of its characteristics, and exemplar test items are created to illustrate it. While it is sometimes appropriate to assign content to a unique level, in most cases the complexity of the stimulus and question determines the level to which a particular test item is assigned.

WorkKeys test specifications for the multiple-choice assessments are unlike the test blueprints used in education. They are not a list of the content topics or objectives to be covered and the number of test items to be assigned to each. Rather, they are more like scoring rubrics used for holistic scoring of constructed-response assessments (White, E. M., 1994). Similarly, the alternatives for a single multiple-choice question may include multiple content classifications, modeling a well-integrated curriculum, yet making the typical approach to test blueprints, which assume that each item measures only one objective, inappropriate.

Prototyping

After development of the general test specifications, ACT test development associates (TDAs) begin writing items for the prototype test. All the items must be written to meet the test specifications and must correspond to the respective skill levels of the test. A number of prototype test items sufficient to create one full-length test form (usually 30 to 40 items) for the skill area are produced.

Each prototype test form (one per skill area) is administered to at least two groups of high school students and two groups of employees. Typically, one group of students and one of employees will be from the same city. The second groups of students and employees will be found in another state with a different situation (for example, if the first groups are from a suburban setting, the second may be from an inner city). The number of examinees varies according to the test format, with more being used for multiple-choice tests than for constructed-response tests. Typically, at least 200 students and 60 employees are divided across the two administration sites for each multiple-choice prototype test form.

During the prototype process, TDAs interview the examinees to gather their reactions to the test instrument, which helps ACT evaluate the functioning of the test specifications. Questions such as whether the prototype items were too hard, too easy, or tested skills outside the realm of the specifications must be answered before development can move to the pretesting stage. Whereas the examinees are asked to provide comments and suggestions about the prototype test form, educators and employers are also invited to review and comment on it. Based on all the information from prototype testing, the test specifications are adjusted if necessary, and additional prototype studies may be conducted. When the prototype process is completed satisfactorily, a written guide for item writers is prepared.

Pretesting

For the pretesting phase, ACT contracts with numerous freelance item writers who produce a large number of items, which ACT staff edit to meet the content, cognitive, and format standards. WorkKeys item writers must be familiar with various work situations and have insight into the use of a particular skill in different employment settings because both content and contextual accuracy are critically important for WorkKeys. A test question containing inaccurate content may be distracting even if the specific content does not affect the examinee's ability to respond correctly to the skills portion of the question. Inaccurate facts, improbable circumstances, or unlikely consequences of a series of procedures or actions are not acceptable. An examinee who knows about a particular workplace should not identify any of the assessment content, circumstances, procedures, or keyed responses as unlikely, inappropriate, or otherwise inaccurate.

Given the wide range of employability skills assessed, verifying content accuracy for WorkKeys is challenging. To help WorkKeys staff detect any possible problems, the item writers write a justification for the best response and for each distractor (incorrect response) for each test item. Both the items and the justifications are checked and, if necessary, the test items are modified.

After the test questions and stimuli have been created and edited, and before administration of the pretesting forms, all items are submitted to external consultants for content and fairness reviews. Qualified experts in the specific skill area being assessed, usually persons using the skills regularly on the job, check for content and contextual accuracy. Members of minority groups review the items to make sure they will not be biased against, or offensive to, racial, ethnic, and gender groups. ACT provides all the reviewers with written guidelines and receives written evaluations back from them.

Table 2.7 shows the numbers of reviewers used for verifying content accuracy and fairness for the current operational assessments. ACT staff respond to every concern the reviewers raise, and any needed adjustments to the test items are made before pretesting.

Table 2.7: Number of Reviewers by Type of Review for the Operational WorkKeys Assessments

	Number of reviewers	
Assessment title	**Content**	**Fairness**
Applied Mathematics	9	8
Reading for Information	13	8

To provide the data required for both classical and item response theory (IRT)–based statistics, each multiple-choice item is administered to a sample of about 2,000 examinees. For practical reasons, most of these examinees are students, although smaller samples of employees are also assessed for each pretest. Then ACT researchers evaluate the psychometric properties (such as reliability and scalability) of each item.

Additionally, statistical, differential item functioning (DIF) analyses of the items are carried out to determine whether items function differently for various groups of individuals (by seeing if responses to items can be correlated with the gender or ethnicity of the examinees). Items that show DIF are eliminated from the item pool. Based on the data collected during pretesting for each skill area, no items in the WorkKeys tests show DIF. Statistical studies can also locate problem items, which are identified during the analysis and are reevaluated by staff and, if necessary, outside experts.

Operational Forms

Pretest item analyses are considered carefully when constructing the forms for operational testing. Alternate and equivalent test forms for each assessment are developed from the pool of items that meet all the content, statistical, and fairness criteria. ACT staff construct at least two equivalent test forms for each assessment. In these forms, both the overall characteristics of the test and the within-level characteristics for content, complexity, and psychometric characteristics are made as similar as possible.

In addition to developing the job-profiling procedure to link the content of the WorkKeys assessments to a specific job, ACT achieves validity through creating well-designed tests. During the development of the assessments, ACT works to minimize the likelihood of adverse impact resulting from use of the WorkKeys tests. Specifically, the assessments are designed to be job-related and fair by ensuring that the items go through a series of screens before they are made available to employers:

- The assessments are criterion-referenced (they use job requirements as the scoring reference, rather than population norms);
- The test specifications are well-defined;
- Items are written by people who have job experience in the workplace and thus the items tap a domain of workplace skill;
- Items measure a particular workplace skill;
- Content and fairness experts review the items to determine possible differences in responses among racial groups and gender; and

- Statistical analyses (for example, differential item functioning) at the item and test level are conducted to monitor the performance of various subgroups.

WorkKeys Assessment Descriptions

Applied Mathematics

The *Applied Mathematics* skill involves the application of mathematical reasoning to work-related problems. The assessment requires the examinee to set up and solve the types of problems and do the types of calculations that actually occur in the workplace. This assessment is designed to be taken with a calculator. As on the job, the calculator serves as a tool for problem solving. A formula sheet that includes, but is not limited to, all formulas required for the assessment is provided. There are five skill levels, with Level 3 requiring the least complex mathematical concepts and calculations and Level 7 requiring the most complex.

Level 3

Problems at Level 3 measure the examinee's skill in performing basic mathematical operations (addition, subtraction, multiplication, and division) and conversions from one form to another, using whole numbers, fractions, decimals, or percentages. Solutions to problems at Level 3 are straightforward, involving a single type of mathematical operation. For example, the examinee might be required to add several numbers or to calculate the correct change in a simple financial transaction.

Level 4

Problems at Level 4 measure the examinee's skill in performing one or two mathematical operations, such as addition, subtraction, or multiplication, on several positive or negative numbers. (Division of negative numbers is not covered until Level 5.) Problems may require adding commonly known fractions, decimals, or percentages (such as ½, .75, 25%), or adding three fractions that share a common denominator. At this level, the examinee is also required to calculate averages, simple ratios, proportions, and rates, using whole numbers and decimals. Problems at this level require the examinee to reorder verbal information before performing calculations. For example, the examinee may be required to calculate sales tax or a sales commission, or to read a simple chart or graph to obtain the information needed to solve a problem.

Level 5

Problems at Level 5 require the examinee to look up and calculate single-step conversions within English or non-English systems of measurement (such as converting from ounces to pounds or from centimeters to meters) or between systems of measurement (such as converting from centimeters to inches). These problems also require calculations using mixed units (such as hours and minutes). Problems at this level contain several steps of logic and calculation. The examinee must determine what information, calculations, and unit conversions are needed to find a solution. For example, the examinee might be asked to calculate perimeters of basic shapes, to calculate percent discounts or mark-ups, or to complete a balance sheet or order form.

Level 6

Problems at Level 6 measure the examinee's skill in using negative numbers, fractions, ratios, percentages, and mixed numbers in calculations. For example, the examinee might be required to calculate multiple rates, to find areas of rectangles or circles and volumes of rectangular solids, or to solve problems that compare production rates and pricing schemes. The examinee might need to transpose a formula before calculating or to look up and use two formulas in conversions within a system of measurement. Level 6 problems may also involve identifying and correcting errors in calculations, and generally require considerable set-up.

Level 7

Problems at Level 7 require multiple steps of logic and calculation. For example, the examinee may be required to convert between systems of measurement that involve fractions, mixed numbers, decimals, or percentages; to calculate multiple areas and volumes of spheres, cylinders, and cones; to set up and manipulate complex ratios and proportions; or to determine the better economic value of several alternatives. Problems may involve more than one unknown, nonlinear functions, and applications of basic statistical concepts (such as error of measurement). The examinee may be required to locate errors in multiple-step calculations. At this level, problem content or format may be unusual, and the information presented may be incomplete or implicit, requiring the examinee to derive the information needed to solve the problem from the setup.

Reading for Information

The *Reading for Information* skill involves reading and understanding work-related instructions and policies. The reading passages and questions in the assessment are based on the actual demands of the workplace. Passages take the form of memos, bulletins, notices, letters, policy manuals, and governmental regulations. Such materials differ from the expository and narrative texts used in most reading instruction, which are usually written to facilitate reading. Workplace communication is not necessarily well-written or targeted to the appropriate audience. Because the *Reading for Information* assessment uses workplace texts, the assessment is more reflective of actual workplace conditions. There are five skill levels, with Level 3 being the least complex and Level 7 the most complex.

Level 3

Questions at Level 3 measure the examinee's skill in reading short, uncomplicated passages that use elementary vocabulary. The reading materials include basic company policies, procedures, and announcements. All of the information needed to answer the questions is stated clearly in the reading materials, and the questions focus on the main points of the passages. At this level, the wording of the questions and answers is similar or identical to the wording used in the reading materials. Questions at Level 3 require the examinee to (1) identify uncomplicated key concepts and simple details; (2) recognize the proper placement of a step in a sequence of events, or the proper time to perform a task; (3) identify the meaning of words that are defined within the passage; (4) identify the meaning of simple words that are not defined within the passage; and (5) recognize the application of instructions given in the passage to situations that are described in the passage.

Level 4

At Level 4, the reading passages are slightly more complex than those at Level 3. They contain more detail and describe procedures that involve a greater number of steps. Some passages describe policies and procedures with a variety of factors that must be considered in order to decide on appropriate behavior. The vocabulary, while elementary, contains words that are more difficult than those at Level 3. For example, the word "immediately" may be used at this level, whereas at Level 3 the phrase "right away" would be used. At this level, the questions and answers are paraphrased from the passage. In addition to the skills tested at the preceding level, questions at Level 4 require the examinee to (1) identify important details that are less obvious than those in Level 3; (2) recognize the application of more complex instructions, some of which involve several steps, to described situations; (3) recognize cause-effect relationships; and (4) determine the meaning of words that are not defined in the reading material.

Level 5

Passages at Level 5 are more detailed, more complicated, and cover broader topics than those at Level 4. Words and phrases may be specialized (for example, jargon and technical terms), and some words may have multiple meanings. Questions at this level typically call for applying information given in the passage to a situation that is not specifically described in the passage. All of the information needed to answer the questions is stated clearly in the passages, but the examinee may need to take several considerations into account in order to choose the correct responses. In addition to the skills tested at the preceding levels, questions at Level 5 require the examinee to (1) identify the paraphrased definition of a technical term or jargon that is defined in the passage; (2) recognize the application of jargon or technical terms to stated situations; (3) recognize the definition of an acronym that is defined in the passage; (4) identify the appropriate definition of a word with multiple meanings; (5) recognize the application of instructions from the passage to new situations that are similar to those described in the reading materials; and (6) recognize the application of more complex instructions to described situations, including conditionals and procedures with multiple steps.

Level 6

Passages at Level 6 are significantly more difficult than those at the previous level. The presentation of the information is more complex; passages may include excerpts from regulatory and legal documents. The procedures and concepts described are more elaborate. Advanced vocabulary, jargon, and technical terms are used. Most information needed to answer the questions correctly is not clearly stated in the passages. The questions at this level require examinees to generalize beyond the stated situation, to recognize implied details, and to recognize the probable rationale behind policies and procedures. In addition to the skills tested at the preceding levels, questions at Level 6 require the examinee to (1) recognize the application of jargon or technical terms to new situations; (2) recognize the application of complex instructions to new situations; (3) recognize, from context, the less common meaning of a word with multiple meanings; (4) generalize from the passage to situations not described in the passage; (5) identify implied details; (6) explain the rationale behind a procedure, policy, or communication; and (7) generalize from the passage to a somewhat similar situation.

Level 7

The questions at Level 7 are similar to those at Level 6 in that they require the examinee to generalize beyond the stated situation, to recognize implied details, and to recognize the probable rationale behind policies and procedures. However, the passages are more difficult: the density of information is higher, the concepts are more complex, and the vocabulary is more difficult. Passages include jargon and technical terms whose definitions must be derived from context. In addition to the skills tested at the preceding levels, questions at Level 7 require the examinee to (1) recognize the definitions of difficult, uncommon jargon or technical terms, based on the context of the reading materials; and (2) figure out the general principles underlying described situations and apply them to situations neither described in nor completely similar to those in the passage.

Technical Characteristics of the WorkKeys Tests

Scoring and Scaling the WorkKeys Tests

The method of assigning level scores to examinees was developed to support two basic assumptions about level scores. First, content experts determined that mastery of a level means being able to correctly answer 80% of the items representing the level. In our method of scoring, the 80% standard is implemented with respect to a pooled (not forms-based) domain of items. This pool of items is referred to here as a “level pool” or “level domain.” For example, in *Applied Mathematics*, each level was represented by 18 items—6 from each of 3 alternate forms. To assess mastery using a level pool, rather than using just the items representing the level on one test form, an item response theory (IRT) model was used, as described below.

The second important assumption about level scores is that an examinee should have mastery of all levels up to and including his or her level score, and nonmastery of higher levels. In WorkKeys job profiling, the level of skill required for a job corresponds to the most complex skill-related tasks a job incumbent would be expected to perform. But the job may also involve less complex skill-related tasks pertaining to lower levels of the same skill. The WorkKeys scoring system must therefore provide reasonable assurance that examinees have a Guttman pattern of mastery over levels, meaning that they have mastery of all levels easier than the level of their score (Guttman, 1950). Since multiple-choice test data contain a significant amount of random error, and there is no formal incorporation of measurement error in Guttman scaling, an IRT model was used for this purpose as well.

The WorkKeys level scoring methods were developed from the data of two or more alternate forms for each skill area. Alternate forms had no items in common, but were designed to be comparable in difficulty. Item statistics from pilot studies were used for this purpose. Five skill levels each were defined for *Applied Mathematics* and *Reading for Information*. For both tests, each level was represented by 6 items on each of three alternate forms. There were thus 30 items per form, a total of 18 items per level, and a grand total of 90 items used to define both the *Applied Mathematics* and *Reading for Information* levels.

Alternate forms for the reading and mathematics skills, as well as for other WorkKeys multiple-choice tests, were administered to randomly equivalent groups of high school juniors and seniors in one state by spiraling forms within classrooms. This data collection process and the analyses that defined the WorkKeys levels are referred to here as the “scaling study.” Summary statistics of number-correct (NC) scores on the *Applied Mathematics* forms used in the scaling study are shown in Table 2.8. The forms are identified here as Forms 1, 2, and 3. Sample sizes ranged from 1,996 to 2,046 per form. The mean NC score ranged from 18.8 to 19.1. Skew and kurtosis were negligible. Reliability coefficients based on the KR_{20} formula ranged from 0.80 to 0.83. Reliability coefficients based on an IRT-method of estimating reliability (Kolen, Zeng & Hanson, 1996; Schulz, Kolen & Nicewander, 1999) were slightly higher (0.82 to 0.85.) It should be noted that these reliability coefficients pertain to the number-correct score, not to the level scores.

The *p*-values of the items constituting the *Applied Mathematics* level pools are displayed in Figure 2.2. This plot shows that item difficulties overlapped across levels but that average item difficulty increased substantially by level (as shown by decreasing mean item p-value). Similar features were exhibited by the *Reading for*

Information test as well as the other multiple-choice WorkKeys tests.

Table 2.8: Statistics and Reliabilities of Number-Correct Scores on *Applied Mathematics* Test Forms

	Form 1	Form 2	Form 3
NC score summary statistics			
Sample size	2,022	2,046	1,996
Mean	18.8	19.0	19.1
SD	5.1	4.9	4.8
Skew	–0.26	–0.38	–0.53
Kurtosis	–0.04	–0.03	0.29
NC score reliability			
KR_{20}	0.83	0.81	0.80
3PL model	0.85	0.83	0.82

The 3-parameter logistic (3PL) model was fit to the data separately for each test form using the computer program BILOG (Mislevy & Bock, 1990). Examinee skill is represented in the 3PL model as a unidimensional, continuous variable, θ (theta). Theta is assumed to be approximately normally distributed in the sample to which the test is administered. Items are represented in the 3PL model by three statistics denoted a, b, and c. These statistics represent, respectively, a, the discriminating power of the item; b, the difficulty of the item; and, c, the lower asymptote of the item response function on theta (θ), which is sometimes referred to as the guessing parameter.

The item statistics from the BILOG analyses were used with the IRT model to predict expected proportion correct (EPC) scores on level pools as a function of θ for each skill. Figure 2.3 shows the EPC score on *Applied Mathematics* level pools as a function of *Applied Mathematics* θ. The curves in this figure are referred to as level response functions. The lower boundary of each *Applied Mathematics* level on the θ scale is shown to be the θ coordinate corresponding to an EPC of 0.8 on the corresponding level pool. For example, the dotted vertical line on the left in Figure 2.3 intersects the Level 3 characteristic curve at coordinates of 0.8 on the EPC axis and –1.43 on the θ axis. This means that an examinee with an *Applied Mathematics* θ of –1.43 has a 0.8 EPC, or an 80% correct true score, on the Level 3 pool of *Applied Mathematics*. The boundary for *Applied Mathematics* Level 3 is thus –1.43.

Figure 2.2: Item *p*-values (*p*) and Mean Item *p*-values (Connected) by Level of Item on WorkKeys *Applied Mathematics* Tests

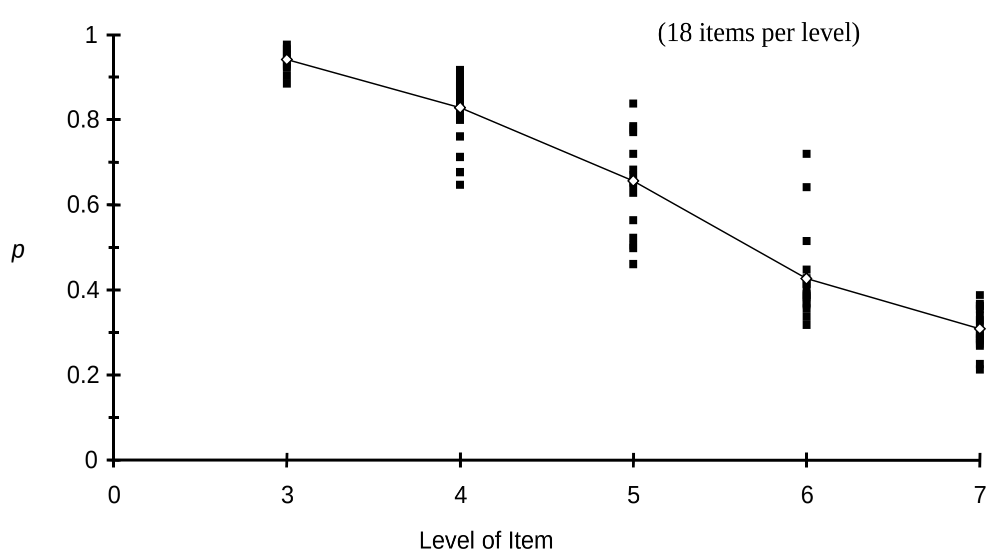
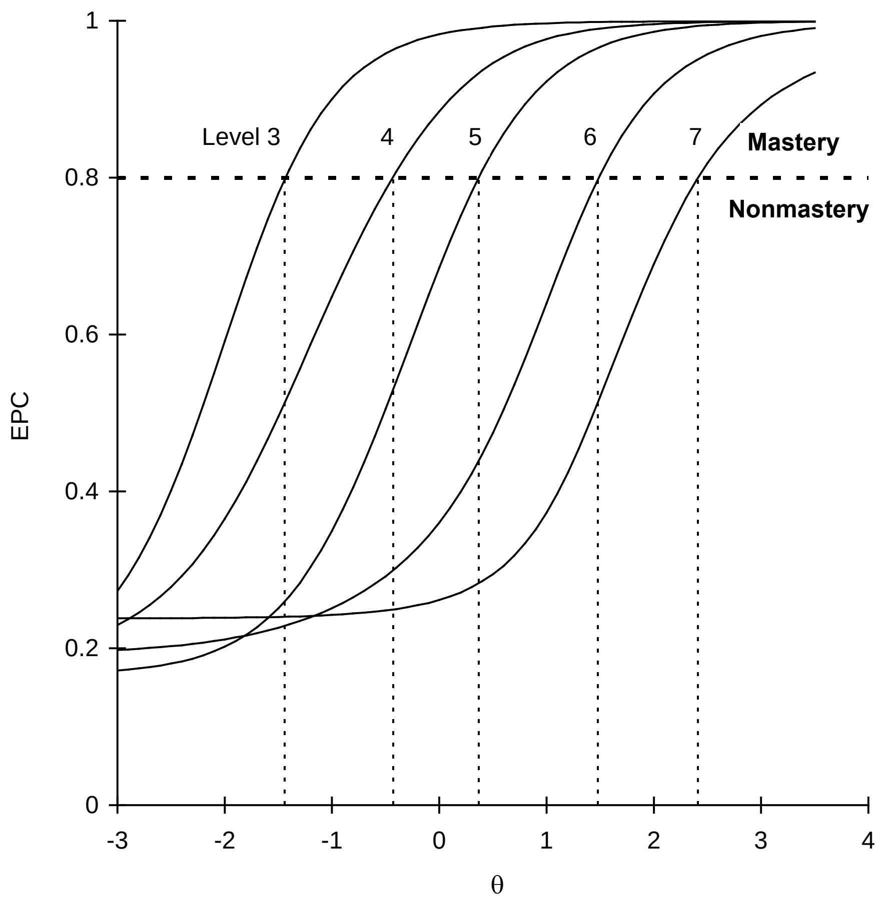


Figure 2.3: ***Applied Mathematics*** **Level Response Functions**



All multiple-choice WorkKeys assessments exhibited level characteristic curves like those in Figure 2.3. The curves were nearly parallel, well spaced, and not overlapping except at low levels associated with guessing. This means that there are substantial differences between adjacent levels of skill and that one can infer a Guttman pattern of level mastery for any examinee: An examinee can be expected to have mastery (that is, ≥ 80% correct) of his or her skill level and all easier levels, but to not have similar mastery of higher levels of skill.

EPC scores represent an examinee's level of skill in two ways that observed scores cannot. First, EPC scores represent performance on a larger set of items than were on any given form. In *Applied Mathematics*, examinees took only 6 items representing a level, but an EPC score represents expected performance on all 18 items representing the level. EPC scores therefore provide a more consistent basis for assigning level scores to examinees who take different forms. Second, EPC scores represent levels of performance that do not necessarily correspond to any observed score. In particular, an 80% correct criterion for mastery does not correspond exactly to an NC score on 6 items (representing a level of *Applied Mathematics* on a form) or 18 items (representing the level more generally).

The EPC method of defining levels of skill rests on the assumptions that the data fit the IRT model and that the samples of examinees taking alternate forms were randomly equivalent. The fit of the data to the model was evaluated by its ability to predict the observed distributions of level scores under three different scoring methods, and to account for observed patterns of mastery over levels (Schulz et al., 1997; Schulz et al., 1999). The fit of the model was judged to be very good in these respects. To estimate the EPC on level pools, item statistics from form-specific BILOG analyses were treated as belonging to a common scale. This treatment rests on the randomly equivalent groups assumption.

Table 2.9 shows the boundary thetas that define levels of WorkKeys skills. The lower boundary of Level 3 on the θ scale for *Applied Mathematics* is shown to be –1.43, as illustrated in Figure 2.3. Similarly, the θ coordinate of the dotted vertical lines representing the lower boundaries of Levels 4, 5, 6, and 7 in Figure 2.3 are shown in the *Applied Mathematics* column of Table 2.9 to be, respectively, –0.43, 0.36, 1.48, and 2.40. Theta values for lower boundaries of other areas of skill were obtained in a similar fashion.

Because the θ distribution in a BILOG analysis is assumed to be standard normal, θ values have approximately the same meaning as Z-scores (standard normal variates). This meaning is useful for understanding how difficult it is to achieve a given level of skill. For example, approximately 8% of a standard normal distribution is below a Z-score of –1.43. It is therefore reasonable to suppose that approximately 8% of the examinees who took the *Applied Mathematics* forms in our scaling study had below Level 3 *Applied Mathematics* skill.

Table 2.9: θ Values at Lower Boundaries of Levels

Level	*Applied Mathematics*	*Reading for Information*
3	–1.43	–1.73
4	–0.43	–0.95
5	0.36	0.06
6	1.48	1.16
7	2.40	–1.73

Table 2.10 shows the range of NC scores assigned to a given level score for each form of *Applied Mathematics* used in the scaling study. For example, on Form 1 of *Applied Mathematics*, NC scores of 12 to 16 were assigned a level score of 3. The cutoff score for a level is the lowest NC score assigned the corresponding level score. The Form 1 cutoff score for Level 3 of *Applied Mathematics* is therefore 12. Similarly, the Form 1 cutoff score for Level 4 is 17.

Table 2.10: Number-Correct Score Ranges by Form and Level of *Applied Mathematics*

Level	Number-correct score range		
	Form 1	Form 2	Form 3
Less than 3	0–11	0–11	0–11
3	12–16	12–16	12–16
4	17–20	17–20	17–20
5	21–24	21–24	21–24
6	25–28	25–27	25–27
7	29+	28+	28+

Table 2.11 shows how cutoff scores were selected. First, the IRT model was used to find a θ for each NC score on each form. The NC score was the true score, rounded to 0.001, for its corresponding θ (Schulz et al., 1999). NC scores whose θ was the closest to the boundary θ for a level were chosen as the cutoff scores for the level.

Table 2.11: Boundary θs and Form-Specific Cutoff θs for Levels of *Applied Mathematics*

Level	Boundary θ	Form-specific cutoff θs		
		Form 1	Form 2	Form 3
3	–1.43	–1.43	–1.51	–1.54
4	–0.43	–0.37	–0.47	–0.49
5	0.36	0.48	0.42	0.40
6	1.48	1.28	1.36	1.36
7	2.40	2.34	2.19	2.56

The θ corresponding to a cutoff score is referred to as a “form-specific cutoff θ.” In Table 2.11, for Level 3 of *Applied Mathematics*, the form-specific cutoff θs were –1.43, –1.51, and –1.54, respectively, for Forms 1, 2, and 3. These θs were associated with an NC score of 12 on their respective forms. Each of these θs was closer to the lower boundary of Level 3 (–1.43) than the θs associated with other NC scores, such as 11 or 13, on their respective forms.

The fact that form-specific cutoff θs do not generally correspond exactly to the boundary θ reflects the difference between continuous and discrete variables. The EPC and θ scales represent achievement and criterion-referenced standards as continuous variables. These scales can represent a 79% or 81% standard of mastery as precisely as an 80% correct standard. NC scores cannot represent most conceivable standards precisely because they are discrete. For example, a 0.8 EPC has no NC representation on an 18-item level pool.

Across-form variation in the θs associated with a particular NC score represents a combination of systematic and random effects across forms. Systematic effects include the true psychometric characteristics of the forms. For example, the fact that the θ associated with a 12 on *Applied Mathematics* Form 3 (–1.54) is lower than the θ associated with a 12 on Form 1 (–1.43) suggests that it may be slightly easier to get a 12 on Form 3 than on Form 1. It is unrealistic to expect no difference between forms. Random effects, however, such as the error in estimates of IRT parameters and random differences in the skill of the Form 1 and Form 3 groups, also play a role.

The cutoff scores for Level 7 of *Applied Mathematics* (Table 2.10) and their associated θs (Table 2.11) illustrate how the selection rule for cutoff scores accommodates differences between forms. The θ for an NC score of 29 on Form 1 (2.34) is lower than the θ for an NC score of 28 on Form 3 (2.56). This result suggests that it is easier to get a score of 29 on Form 1 than it is to get a score of 28 on Form 3. This difference cannot help but lead to different cutoff scores for a level whose boundary θ is in between these two values. Each value is closest to the Level 7 boundary (2.40) within its respective form. The Form 1 cutoff score (29) is therefore one point higher than the Form 3 cutoff score (28).

From these examples, it is clear that the psychometric differences between test forms may be too complex to permit simple statements such as "Form 1 is easier than Form 3." The examples suggest that it is harder to get a score of 12 on Form 1 than on Form 3, but easier to get a score of 29 on Form 1 than a score of 28 on Form 3. These differences may be explained by between-form differences in the distributions of the item statistics. It is not necessary to determine the reasons for these differences, however, to take them into account when selecting cutoff scores.

Given that cutoff scores were selected in this way, it is remarkable that cutoff scores were so often the same across forms. With the exception of the Form 1 cutoff score for Level 7 (29), the cutoff scores for levels of *Applied Mathematics* were the same across all three forms—12 for Level 3, 17 for Level 4, 21 for Level 5, 25 for Level 6, and 28 for Level 7. These results attest to the reliability of item statistics from pilot data and to the care with which these statistics were used to make the alternate forms psychometrically equivalent.

Since the forms were administered to randomly equivalent groups, and cutoff scores were selected to implement standards consistently across forms, the distributions of level scores should be similar across forms. Table 2.12 shows results pertaining to this expectation. The percentage at each level of *Applied Mathematics*, rounded to the nearest whole number, is shown by form. The mean and standard deviation of level scores is also shown by form. "Below 3" level scores were coded as "2" to compute the mean and standard deviation. The distributions of level scores are similar across forms. Means and standard deviations differ by no more than 0.1. The percentages at a given level differ by no more than 4 points. In particular, the percentage of Level 7 scores was 2, 3, and 2, respectively, for Forms 1, 2, and 3. From the similarity of these percentages, we concluded that a cutoff score of 29 for Level 7 on Form 1 was not too high in comparison to a cutoff score of 28 on the other two forms.

Table 2.12: Summary Statistics of Level Scores by Form of *Applied Mathematics*

	Percentage		
Level	**Form 1**	**Form 2**	**Form 3**
Below 3	8	8	7
3	22	20	20
4	31	32	32
5	25	29	29
6	11	9	11
7	2	3	2
Mean level score:	4.1	4.2	4.2
Standard deviation	1.2	1.2	1.1

Cutoff scores for alternate forms of all multiple-choice tested WorkKeys skills were obtained as described here for *Applied Mathematics*. Results for the other skills were similar to those presented here. Cutoff scores were equal across forms in most cases, and the resulting distributions of level scores were similar across forms. Form-specific results for the other skills are not shown here because the purpose of this chapter is to provide a general illustration of how level scores were obtained from NC scores. Form-specific results for *Applied Mathematics* show how the method performed generally.

The method of selecting WorkKeys cutoff scores is slightly lenient. The cutoff θ does not necessarily exceed the boundary theta. For example, the Level 3 cutoff θ for Form 2 of *Applied Mathematics*, –1.51, does not exceed the Level 3 boundary θ of –1.43. This practice tends to produce a higher false-positive–to–false-negative error ratio and to produce a higher overall classification error rate than if the cutoff θ exceeded the boundary θ.

A slightly lenient scoring rule was deliberately chosen for two important reasons. First, the current scoring procedure replaces one that was also lenient (Schulz et al., 1997; Schulz et al., 1999). The current procedure and the previous scoring procedure produce similar frequency distributions of observed level scores. This is important for connecting current results with past results for WorkKeys users.

Second, a lenient implementation of the 0.8 EPC standard in WorkKeys is justified by the error inherent in measuring with reference to a standard. In addition to the measurement error associated with an examinee's score, there is also error in setting a criterion-referenced standard. One or both of these types of error are typically cited in choosing a cutoff score that is more lenient, and gives the benefit of doubt to the examinee. Leniency typically takes the form of a cutoff score that is one or more standard errors of measurement below the score that strictly represents the standard. Our particular method of scoring WorkKeys tests is less lenient than this. Strict implementation of the 0.8 EPC standard would require the cutoff θ to exceed the boundary θ. In about half the cases, it already does. In the other half, the cutoff score would be only one point higher. Thus, about half the time, the cutoff score is only one NC point lower than a strict implementation of the standard would require. One NC point is less than one standard error of measurement on the NC scale for the WorkKeys tests.

Reliability, Classification Consistency, Classification Error of the WorkKeys Tests

Test publishers are advised to provide indices that reflect random effects on test scores (AERA, 1999). The indices provided in this chapter fall into three broad categories: (1) reliability and standard error, (2) classification consistency, and (3) classification error.

One definition of reliability is "the correlation between two parallel forms of a test" (Gulliksen, 1987, p. 13). In the theory for this definition, the observed score of a given examinee *i*, x_i, is a chance variable with an unknown distribution. The mean, μ_i, and standard deviation, σ_i, of this distribution are called the "examinee's true score" and "standard error of measurement," respectively. The standard error of measurement generally varies with the true score, and is not the same for every examinee. The reliability of the observed score, *X*, for a group of examinees is related to the standard errors of examinees' scores through the equation:

$$\rho = 1 - \frac{\sigma_e^2}{\sigma_X^2},$$

where ρ is the reliability, σ_e^2 is the mean squared measurement error over examinees, and σ_X^2 is the variance of *X* over examinees. The mean squared measurement error can be as great as σ_X^2 or as small as 0.

These extreme values correspond to the limits of reliability which are, respectively, 0 and 1. A reliability coefficient of 1 means that there is no measurement error for any examinee—that each examinee would earn the same score on every parallel test.

Unfortunately, reliability coefficients and standard errors have limited meaning for WorkKeys tests. WorkKeys tests are primarily classification tests. They are designed to permit accurate at-or-above classifications of examinees with regard to the particular level of skill that may be required in a given job or setting. Professional standards for testing advise publishers of classification tests to provide information about the percentage of examinees that would be classified in the same way on two applications of the same form or alternate forms (AERA, 1999). These standards note that reliability coefficients and standard errors do not directly answer this practical question.

Also, as criterion-referenced classification tests, WorkKeys level scores are not defined primarily to represent differences between examinees. Only five criterion-referenced levels are defined for *Reading for Information* and *Applied Mathematics* WorkKeys tests. These levels are labeled with successive integers (3, 4, 5, 6, and 7) for convenience. These integers do not imply that differences between levels are in any sense comparable or equal. The meaning, as well as the specific values, of reliability coefficients and standard errors depends on the score scale and changes with the meaning of differences between scores. Reliability coefficients tend to be lower and standard errors of measurement higher as the number of score scale points decreases. In particular, the reliability of level scores is lower than the reliability of NC scores on WorkKeys tests (for example, compare 3PL model NC reliabilities in Table 2.8 with the reliability of level scores reported in Table 2.13 for *Applied Mathematics*). Since only level scores are reported for WorkKeys tests in general, the reliability and standard error of only level scores are reported in this chapter. No reliability coefficient, however, bears directly on how random error affects the classification function of WorkKeys tests.

Indices of classification consistency are more directly informative about the effects of measurement error on a classification test. Classification consistency is defined here as "the proportion or percentage of examinees who would be classified the same way by two parallel tests." As a proportion, classification consistency has the same range as the reliability coefficient: 0 to 1, with 1 being the maximum or best possible. As a percentage, classification consistency ranges from 0 to 100.

Indices of classification error provide additional information about the effects of measurement error on a classification test. Two types of classification errors are defined in this chapter. A “false positive” error occurs when an examinee is classified into a level or range of levels that is higher than his or her true level. A “false negative” error occurs when an examinee is classified into a level or range of levels that is lower than his or her true level. Total classification error is the sum of these two types of errors. The total error rate ranges from 0 to 1, with 0 being the best possible result.

Estimates of classification error are critical and perhaps more important than estimates of classification consistency for evaluating a classification test. Most users would consider a less consistent test to be better than a more consistent one if it has a lower classification error rate.

Estimates of reliability, classification consistency, and classification error were derived from a scaling study and pilot data (described on page 20) using the IRT methodology described in Schulz, Kolen & Nicewander (1997, 1999). This methodology performed well when compared with classical methods (Lee, Brennan & Hanson, 2000). Results for each skill (*Applied Mathematics* and *Reading for Information*) have been averaged over two or more alternate forms. This does not mean that the indices reported here represent test-retest effects or even differences across randomly parallel forms. The IRT-based estimates represent only the random error in a single test form, or differences across strictly parallel forms (Yen, 1983). All of the indices reported in this section are affected by the distribution of skill in the scaling and pilot studies.

The upper panel of Table 2.13 shows the actual or predicted percentages of students in the scaling or pilot studies who scored at each level of a given skill. For example, 21% of the examinees in the scaling study earned a level score of 3 in *Applied Mathematics*, and 32% earned a level score of 4. Percentages above 0.5 are rounded to the nearest integer. Percentages less than 0.5 are rounded to the nearest 0.1. Because of rounding, percentages within columns may not add to 100.

All of the percentages in the upper panel of Table 2.13 show the actual percentages of level scores in the scaling study. Level percentages were predicted by applying the IRT model to item statistics from the pilot studies for this test and by assuming a standard normal θ distribution, but these are not shown in Table 2.13. However, the predicted percentages were very close to the actual percentages shown in Table 2.13. The equivalence of IRT-predicted percentages and actual percentages is one indication that the IRT model fit the WorkKeys data well enough to predict reliability, classification consistency, and classification error (Schulz et al., 1997, 1999; see also Lee, Brennan & Hanson, 2000).

Table 2.13: Frequency Distributions[a] and Reliability of Level Scores of WorkKeys Multiple-Choice Tests

Level	*Applied Mathematics*	*Reading for Information*
Below 3	8	6
3	21	8
4	32	38
5	27	30
6	10	17
7	3	2
Mean	4.2	4.5
Standard deviation	1.2	1.1
Standard error	0.55	0.59
Reliability	0.78	0.72

[a] Frequencies are reported as percentages. Because of rounding, percentages within columns may not add to 100.

The bottom panel of Table 2.13 shows the summary statistics corresponding to percentages in the upper panel. These include the mean and standard deviation of level scores earned by students in the scaling study, the root mean squared error (standard error), and the reliability of the level scores. *Applied Mathematics* levels scores had a mean of 4.2, and a standard deviation of 1.2. Estimates of the standard error and reliability of *Applied Mathematics* level scores were, respectively, 0.55 and 0.78. To compute these statistics, a level score of 2 was assigned to examinees who scored below Level 3.

Table 2.14 shows estimates of classification consistency for each skill. The first row, labeled “Exact,” shows the percentage of examinees in the scaling study who would receive the same level score from two strictly parallel test forms. For example, if an examinee were to take two strictly parallel forms of *Applied Mathematics* and score a Level 3 on both forms, this would be a case of exact agreement. For *Applied Mathematics*, we estimated that such cases would amount to 52% of the examinees in the scaling study.

The remaining rows in Table 2.14 show the consistency of at-or-above classifications separately by level. Entries in the row labeled “≥5,” for example, reflect the consistency of classifying examinees with respect to being at or above level 5. If an examinee were to take two strictly parallel forms of *Applied Mathematics* and receive a level score of 4 the first time and 5 the second, he or she would not be consistently classified with respect to being at or above Level 5 (≥5), but would be consistently classified with respect to being at or above any other level. For example, both a 4 and a 5 are at or above Level 4 (≥4) and both are below Level 6 (which corresponds to the ≥6 type of classification).

Classification consistency is clearly higher for at-or-above classifications than for exact classifications. At-or-above consistency of *Applied Mathematics* scores are estimated to be not less than 81% (for ≥5), and is as high as 97% (for ≥7).

Table 2.14: Predicted Classification Consistency

Type of classification[a]	*Applied Mathematics*	*Reading for Information*
Exact	52	50
≥ 3	94	96
≥ 4	84	90
≥ 5	81	78
≥ 6	91	84
≥ 7	97	96

[a] Exact classifications specify a specific skill level for the examinee; ≥ classifications specify whether the examinee is at or above the indicated level.

Table 2.15 shows the estimated percentages of false positive, false negatives, and total classification error for each skill. These percentages are again reported separately for two types of classification: exact and at-or-above. A score of Level 5 for an examinee whose true level is 4 is a false-positive error in an “Exact” classification, because 5 is higher than 4. This case is also a false positive error with respect to being at or above Level 5, because the 5 would place the examinee in a higher score range (≥5) than the true score (4) merits. This case represents no error with respect to the other at-or-above classifications, however, because none of them would place a 4 in a different category than a 5. For example, a 4 and a 5 are both at or above Level 3 (≥3), and both below Level 6 (corresponding to the “≥6” row/type of classification).

According to the values in the “Exact” row of Table 2.15, 23% of the examinees in the scaling study who took *Applied Mathematics* forms received a level score that was too high (false positive). Another 14% received a level score that was too low (false negative), given their true level of skill in *Applied Mathematics*. The percentage shown in the “Total” column for “Exact” type of classifications in Table 2.15 is the sum of the percentages of false negative and false positive classification errors—38% in this example. Because of rounding, the percentages shown may not add up exactly.

The predicted error percentages for at-or-above classifications are lower than those for exact classifications. For *Applied Mathematics*, the maximum total error rate for any at-or-above classification is only 13% (for ≥5) and the lowest is only 2% (for ≥7).

Table 2.15: Predicted Classification Error[a]

Type of classification[b]	*Applied Mathematics*			*Reading for Information*		
	False +	False –	Total	False +	False –	Total
Exact	23	14	38	27	13	40
≥ 3	2	2	4	1	2	3
≥ 4	6	6	12	4	3	8
≥ 5	7	6	13	10	6	16
≥ 6	7	1	7	10	2	12
≥ 7	2	0	2	3	.01	3

[a] Reported as percentage of examinees in scaling study.
[b] Exact classifications specify a specific skill level for the examinee; “≥” classifications specify whether the examinee is at or above the indicated level.

Estimates of classification error and consistency are sensitive to the distribution of skill in the scaling study. For example, the lower boundary on the θ scale for Level 5 of *Applied Mathematics*, 0.36 (see Table 2.9), is near the zero-mean of the *Applied Mathematics* θ distribution used to compute classification consistency and classification error. This means that the true skill of a relatively large proportion of these examinees was close to the Level 5 boundary. Generally, the closer an examinee's true skill is to a criterion, the more likely he or she is to be misclassified because of measurement error. Given this fact, an 81% classification consistency and a 13% total classification error rate for ≥5 *Applied Mathematics* classifications seems very good.

By the same reasoning, however, a 97% classification consistency and a 2% total classification error rate for ≥7 classifications in *Applied Mathematics* are probably overly optimistic estimates. The Level 7 boundary for *Applied Mathematics*, 2.40 (see Table 2.9), is far above the skill of most examinees in a standard normal θ distribution. Applicants for Level 7 jobs, however, will probably have skill closer to the Level 7 boundary. In that case, the classification consistency would be lower, and classification error higher, than the values in Tables 2.14 and 2.15 indicate.

Validation Issues

The WorkKeys assessments are designed for use by business and education. Two of the most frequent business uses of WorkKeys are screening job applicants by verifying that they have the basic skill levels required to perform the job and identifying skill gaps among employees to determine what basic skills training is needed and by whom. In general, the use of WorkKeys in educational settings and employment training is less prone to legal ramifications than the use of the assessments for selecting and promoting employees. Consult the *WorkKeys Applied Mathematics Technical Manual* (ACT, 2008a) and the *WorkKeys Reading for Information Technical Manual* (ACT, 2008b) for additional information.

Score Distributions of the WorkKeys Assessments

An important aspect of a technical handbook for an assessment instrument is a comprehensive description of the assessment score distributions. For norm-referenced instruments, this usually involves presenting a table of means and standard deviations or standard errors of the scores from the sample used to establish norms.

The WorkKeys assessments are, by design, criterion-referenced instruments, so no national study has been conducted to establish any norms. It is, however, necessary to provide WorkKeys assessment users with information about the characteristics of the WorkKeys assessment score distributions. Also, even though the same secure assessments may be used over the years, the test-takers, as a group, change over time. Therefore, the information about the score distributions should be updated periodically. This section provides detailed information about the score distribution characteristics of a sample of examinees who took WorkKeys assessments in fall 2009 and spring 2010.

Unlike norm-referenced assessment scores, the WorkKeys assessments use only five level score points in the reporting scale. These level scores are ordinal in nature as they form a hierarchy. Therefore, it is not useful or meaningful to describe the score distributions with means, standard deviations, or standard errors. Instead, numbers and percentages of the examinees in the sample at each skill level are used to report the score distributions of the sample in this section.

Table 2.16 contains the numbers and percentages of the examinees who scored at each level of each operational WorkKeys assessment. These statistics are provided for information only and do not constitute any norms, nor should they be used as such for the WorkKeys assessments.

Table 2.16: Numbers and Percentages of Examinees Who Scored at Each Level (Based on 2010–2011 Data)

	Applied Mathematics		*Reading for Information*	
Level	**Number**	**Percent**	**Number**	**Percent**
<3	54,124	7.0	56,309	7.3
3	153,424	19.9	149,834	19.4
4	220,301	28.5	279,520	36.3
5	168,928	21.9	225,729	29.3
6	126,528	16.4	32,934	4.3
7	49,021	6.3	26,646	3.5
Total	772,326		770,972	

Interpretation of WorkKeys Scores

Interpretation of WorkKeys scores with respect to education and training revolves around what the individual can and cannot do within any given skill area. However, there needs to be some standard by which to judge how much of a skill an individual needs. It is important to remember that interpretation of scores can be accomplished with respect to the content of the skill and the resultant level achieved by an individual. This works well when dealing with educational or training institutions. Scores may also be interpreted with respect to requirements of the world of work in the form of skill requirements for specific jobs or for more general occupational clusters or job families. Training institutions can set a minimum competency standard specifying that all individuals must attain a specific level of skill before they exit a program. However, this standard may be too high or too low for some individuals when compared with what is needed in their chosen fields. It is also possible to compare each individual with a standard that relates to his or her job choice or future educational plans. The occupational profiles collected by ACT are examples of such standards. For additional information, please consult **www.act.org/workkeys/index.html**.

The ACT Test

The ACT Test Program is a comprehensive system of data collection, processing, and reporting designed to help high school students develop postsecondary educational plans and to help postsecondary educational institutions meet the needs of their students. One component of the ACT Test Program is the ACT Test, a battery of four multiple-choice tests—English, Mathematics, Reading, and Science—and a Writing Test. The ACT Test Program also includes an interest inventory, and it collects information about students' high school courses and grades, educational and career aspirations, extracurricular activities, and special educational needs. The ACT is taken under standardized conditions; the other noncognitive components are completed during an in-school session on a day before the Day 1 administration of the PSAE.

ACT Test data are used for many purposes. High schools use ACT data in academic advising and counseling, evaluation studies, accreditation documentation, and public relations. Colleges use ACT results for admissions and course placement. States use the ACT Test as part of their statewide assessment systems. Many of the agencies that provide scholarships, loans, and other types of financial assistance to students tie such assistance to students' academic qualifications. Many state and national agencies also use ACT data to identify talented students and award scholarships.

Philosophical Basis for the ACT

Underlying the ACT tests of educational achievement is the belief that students' preparation for college is best assessed by measuring, as directly as possible, the academic skills that they will need to perform college-level work. The required academic skills can be assessed most directly by reproducing as faithfully as possible the complexity of college-level work. Therefore, the tests of educational achievement are designed to determine how skillfully students solve problems, grasp implied meanings, draw inferences, evaluate ideas, and make judgments in content areas important to success in college.

Accordingly, the tests of educational achievement are oriented toward the general content areas of college and high school instructional programs. The test questions require students to integrate the knowledge and skills they possess in major curriculum areas with the information provided by the test. Thus, scores on the tests have a direct and obvious relationship to the students' educational achievement in curriculum-related areas and possess a meaning that is readily grasped by students, parents, and educators.

Tests of general educational achievement are used in the ACT because, in contrast to other types of tests, they best satisfy the diverse requirements of tests used to facilitate the transition from secondary to postsecondary education. By comparison, measures of examinee knowledge of specific course content (as opposed to curriculum areas) do not readily provide a common baseline for comparing students for the purposes of admission, placement, or awarding scholarships because high school courses vary extensively. In addition, such tests might not measure students' skills in problem solving and in the integration of knowledge from a variety of courses.

Tests of educational achievement can also be contrasted with tests of academic aptitude. The stimuli and test questions for aptitude tests are often chosen precisely for their dissimilarity to instructional materials, and each test within a battery of aptitude tests is designed to be homogeneous in psychological structure. With such an approach, these tests may not reflect the complexity of college-level work or the interactions among the skills measured. Moreover, because aptitude tests are not directly related to instruction, they may not be as useful as tests of educational achievement for making placement decisions in college.

The advantage of tests of educational achievement over other types of tests for use in the transition from high school to college becomes evident when their use is considered in the context of the educational system. Because tests of education achievement measure many of the same skills that are taught in high school, the best preparation for tests of educational achievement is high school course work. Long-term learning in school, rather than short-term cramming and coaching, becomes the best form of test preparation. Thus, tests of educational achievement tend to serve as motivators by sending students a clear message that high test scores are not simply a matter of innate ability but reflect a level of achievement that has been earned as a result of hard work.

Because the ACT stresses such general concerns as the complexity of college-level work and the integration of knowledge from a variety of sources, students may be influenced to acquire skills necessary to handle these concerns. In this way, the ACT may serve to aid high schools in developing in their students the higher-order thinking skills that are important for success in college and later life.

The tests of the ACT therefore are designed not only to accurately reflect educational goals that are widely accepted and judged by educators to be important, but also to give educational considerations, rather than statistical and empirical techniques, paramount importance.

Description of the ACT

The ACT contains four multiple-choice tests—English, Mathematics, Reading, and Science. These tests are designed to measure skills that are most important for success in postsecondary education and that are acquired in secondary education.

The fundamental idea underlying the development and use of these tests is that the best way to determine how well prepared students are for further education is to measure as directly as possible the academic skills that students will need to perform college-level work. The content specifications describing the knowledge and skills to be measured by the ACT were determined through a detailed analysis of relevant information: First, the curriculum frameworks for grades seven through twelve were obtained for all states in the United States that had published such frameworks. Second, textbooks on state-approved lists for courses in grades seven through twelve were reviewed. Third, educators at the secondary and postsecondary levels were consulted on the importance of the knowledge and skills included in the reviewed frameworks and textbooks.

Because one of the primary purposes of the ACT is to assist in college admission decisions, in addition to taking the steps described above, ACT conducted a detailed survey to ensure the appropriateness of the content of the ACT tests for this particular use. College faculty members across the nation who were familiar with the academic skills required for successful college performance in language arts, mathematics, and science were surveyed. They were asked to rate numerous knowledge and skill areas on the basis of their importance to success in entry-level college courses and to indicate which of these areas students should be expected to master before entering the most common entry-level courses. They were also asked to identify the knowledge and skills whose mastery would qualify a student for advanced placement. A series of consultant panels were convened, at which the experts reached consensus regarding the important knowledge and skills in English and reading, mathematics, and science, given current and expected curricular trends.

Curriculum study is ongoing at ACT. Curricula in each content area (English, reading, mathematics, and science) in the ACT tests are reviewed on a periodic basis. ACT's analyses include reviews of tests, curriculum guides, and national standards; surveys of current instructional practice (ACT, 2009); and meetings with content experts.

The tests in the ACT are designed to be developmentally and conceptually linked to those of EXPLORE (Grades 8 and 9) and PLAN (Grade 10). To reflect that continuity, the names of the content area tests are the same across the three programs. Moreover, the programs are similar in their focus on thinking skills and in their common curriculum base. The test specifications for the ACT are consistent with, and should be seen as a logical extension of, the content and skills measured in EXPLORE and PLAN.

The English Test

The ACT English Test is a 75-item, 45-minute test that measures understanding of the conventions of standard written English (punctuation, grammar and usage, and sentence structure) and of rhetorical skills (strategy, organization, and style). Spelling, vocabulary, and rote recall of rules of grammar are not tested. The test consists of five prose passages, each accompanied by a sequence of multiple-choice test items. Different passage types are employed to provide a variety of rhetorical situations. Passages are chosen not only for their appropriateness in assessing writing skills, but also to reflect students' interests and experiences. Most items refer to underlined portions of the passage and offer several alternatives to the portion underlined. These items include "NO CHANGE" to the underlined portion in the passage as one of the possible responses. Some items are identified by a number or numbers in a box. These items ask about a section of the passage, or about the passage as a whole. The student must decide which choice is most appropriate in the context of the passage, or which choice best answers the question posed.

Three scores are reported for the English Test: a total test score based on all 75 items, a subscore in Usage/Mechanics based on 40 items, and a subscore in Rhetorical Skills based on 35 items.

The Mathematics Test

The ACT Mathematics Test is a 60-item, 60-minute test that is designed to assess the mathematical reasoning skills that students across the United States have typically acquired in courses taken up to the beginning of Grade 12. The test presents multiple-choice items that require students to use their mathematical reasoning skills to solve practical problems in mathematics. Knowledge of basic formulas and computational skills are assumed as background for the problems, but memorization of complex formulas and extensive computation are not required. The material covered on the test emphasizes the major content areas that are prerequisite to successful performance in entry-level courses in college mathematics. Six content areas are included: pre-algebra, elementary algebra, intermediate algebra, coordinate geometry, plane geometry, and trigonometry.

The items included in the Mathematics Test cover four cognitive levels: knowledge and skills, direct application, understanding concepts, and integrating conceptual understanding. "Knowledge and skills" items require the student to use one or more facts, definitions, formulas, or procedures to solve problems that are presented in purely mathematical terms. "Direct application" items require the student to use one or more facts, definitions, formulas, or procedures to solve straightforward problem sets in real-world situations. "Understanding concepts" items test the student's depth of understanding of major concepts by requiring reasoning from a concept to reach an inference or a conclusion. "Integrating conceptual understanding" items test the student's ability to achieve an integrated understanding of two or more major concepts so as to solve nonroutine problems.

Calculators, although not required, are permitted for use on the Mathematics Test. Almost any four-function, scientific, or graphing calculator may be used on the Mathematics Test. A few restrictions do apply to the calculator used. These restrictions can be found in the current year's *ACT User Handbook* or on ACT's website at **www.act.org**.

Four scores are reported for the Mathematics Test: a total test score based on all 60 items, a subscore in Pre-Algebra/Elementary Algebra based on 24 items, a subscore in Intermediate Algebra/Coordinate Geometry based on 18 items, and a subscore in Plane Geometry/Trigonometry based on 18 items.

The Reading Test

The ACT Reading Test is a 40-item, 35-minute test that measures reading comprehension as a product of skill in referring and reasoning. That is, the test items require students to derive meaning from several texts by: (1) referring to what is explicitly stated and (2) reasoning to determine implicit meanings. Specifically, items ask students to use referring and reasoning skills to determine main ideas; locate and interpret significant details; understand sequences of events; make comparisons; comprehend cause-effect relationships; determine the meaning of context-dependent words, phrases, and statements; draw generalizations; and analyze the author's or narrator's voice or method. The test comprises four prose passages that are representative of the level and kinds of text commonly encountered in first-year college curricula; passages on topics in the social sciences, the natural sciences, prose fiction, and the humanities are included. Each passage is preceded by a heading that identifies what type of passage it is (e.g., "Prose Fiction"), names the author, and may include a brief note that helps in understanding the passage. Each passage is accompanied by a set of multiple-choice test items. These items focus on the complex of complementary and mutually supportive skills that readers must bring to bear in studying written materials across a range of subject areas. They do not test the rote recall of facts from outside the passage or rules of formal logic, nor do they contain isolated vocabulary questions.

Three scores are reported for the Reading Test: a total test score based on all 40 items, a subscore in Social Studies/Sciences reading skills (based on the 20 items in the social sciences and natural sciences sections of the test), and a subscore in Arts/Literature reading skills (based on the 20 items in the prose fiction and humanities sections of the test).

The Science Test

The ACT Science Test is a 40-item, 35-minute test that measures the interpretation, analysis, evaluation, reasoning, and problem-solving skills required in the natural sciences. The content of the Science Test is drawn from biology, chemistry, physics, and the Earth/space sciences, all of which are represented in the test. Students are assumed to have a minimum of two years of introductory science, which ACT's National Curriculum Studies have identified as typically one year of biology and one year of physical science and/or Earth science. Thus, it is expected that students have acquired the introductory content of biology, physical science, and Earth science, are familiar with the nature of scientific inquiry, and have been exposed to laboratory investigation.

The test presents seven sets of scientific information, each followed by a number of multiple-choice test items. The scientific information is conveyed in one of three different formats: data representation (graphs, tables, and other schematic forms), research summaries (descriptions of several related experiments), or conflicting viewpoints (expressions of several related hypotheses or views that are inconsistent with one another).

The items included in the Science Test cover three cognitive levels: understanding, analysis, and generalization. "Understanding" items require students to recognize and understand the basic features of, and concepts related to, the provided information. "Analysis" items require students to examine critically the relationships between the information provided and the conclusions drawn or hypotheses developed. "Generalization" items require students to generalize from given information to gain new information, draw conclusions, or make predictions.

One score is reported for the Science Test: a total test score based on all 40 items.

Test Development Procedures for the ACT Multiple-Choice Tests

This section describes the procedures that are used in developing the four multiple-choice tests described above. The test development cycle required to produce each new form of the ACT tests takes as long as two and one-half years and involves several stages, beginning with a review of the test specifications.

Reviewing Test Specifications

Two types of test specifications are used in developing the ACT tests: content specifications and statistical specifications.

Content specifications

Content specifications for the ACT tests were developed through the curricular analysis discussed above. While care is taken to ensure that the basic structure of the ACT tests remains the same from year to year so that the scale scores are comparable, the specific characteristics of the test items used in each specification category are reviewed regularly. Consultant panels are convened to review both the tryout versions and the new forms of each test to verify their content accuracy and the match of the content of the tests to the content specifications. At these panels, the characteristics of the items that fulfill the content specifications are also reviewed. While the general content of the test remains constant, the particular kinds of items in a specification category may change slightly. The basic structure of the content specifications for each of the ACT multiple-choice tests is provided in Tables 2.17–2.20.

Statistical specifications

Statistical specifications for the tests indicate the level of difficulty (proportion correct) and minimum acceptable level of discrimination (biserial correlation) of the test items to be used.

The tests are constructed with a target mean item difficulty of about 0.58 for the ACT population and a range of difficulties from about 0.20 to 0.89. The distribution of item difficulties was selected so that the tests will effectively differentiate among students who vary widely in their level of achievement.

With respect to discrimination indices, items should have a biserial correlation of 0.20 or higher with test scores measuring comparable content. Thus, for example, performance on mathematics items should correlate 0.20 or higher with performance on the relevant Mathematics Test subscore.

Table 2.17: Content Specifications for the ACT English Test

Six elements of effective writing are included in the English Test. These elements and the approximate proportion of the test devoted to each are given in the table.

Content/Skills	Proportion of test		Number of items	
Usage/Mechanics	0.53			40
Punctuation[a]		0.13	10	
Grammar and Usage[b]		0.16	12	
Sentence Structure[c]		0.24	18	
Rhetorical Skills	0.47			35
Strategy[d]		0.16	12	
Organization[e]		0.15	11	
Style[f]		0.16	12	
Total	**1.00**			**75**

Scores reported: Usage/Mechanics
Rhetorical Skills
Total test score

[a]*Punctuation.* The items in this category test the student's knowledge of the conventions of internal and end-of-sentence punctuation, with emphasis on the relationship of punctuation to meaning (for example, avoiding ambiguity, indicating appositives).

[b]*Grammar and Usage.* The items in this category test the student's understanding of agreement between subject and verb, between pronoun and antecedent, and between modifiers and the words modified; verb formation; pronoun case; formation of comparative and superlative adjectives and adverbs; and idiomatic usage.

[c]*Sentence Structure.* The items in this category test the student's understanding of relationships between and among clauses, placement of modifiers, and shifts in construction.

[d]*Strategy.* The items in this category test the student's ability to develop a given topic by choosing expressions appropriate to an essay's audience and purpose; to judge the effect of adding, revising, or deleting supporting material; and to judge the relevancy of statements in context.

[e]*Organization.* The items in this category test the student's ability to organize ideas and to choose effective opening, transitional, and closing sentences.

[f]*Style.* The items in this category test the student's ability to select precise and appropriate words and images, to maintain the level of style and tone in an essay, to manage sentence elements for rhetorical effectiveness, and to avoid ambiguous pronoun references, wordiness, and redundancy.

Table 2.18: Content Specifications for the ACT Mathematics Test

The items in the Mathematics Test are classified with respect to six content areas. These areas and the approximate proportion of the test devoted to each are given in the table.

Content Area	Proportion of test	Number of items
Pre-Algebra[a]	0.23	14
Elementary Algebra[b]	0.17	10
Intermediate Algebra[c]	0.15	9
Coordinate Geometry[d]	0.15	9
Plane Geometry[e]	0.23	14
Trigonometry [f]	0.07	4
Total	**1.00**	**60**

Scores reported: Pre-Algebra/Elementary Algebra
Intermediate Algebra/Coordinate Geometry
Plane Geometry/Trigonometry
Total test score

[a]*Pre-Algebra.* Items in this content area are based on operations using whole numbers, decimals, fractions, and integers; place value; square roots and approximations; the concept of exponents; scientific notation; factors; ratio, proportion, and percent; linear equations in one variable; absolute value and ordering numbers by value; elementary counting techniques and simple probability; data collection, representation, and interpretation; and understanding simple descriptive statistics.

[b]*Elementary Algebra.* Items in this content area are based on properties of exponents and square roots, evaluation of algebraic expressions through substitution, using variables to express functional relationships, understanding algebraic operations, and the solution of quadratic equations by factoring.

[c]*Intermediate Algebra.* Items in this content area are based on an understanding of the quadratic formula, rational and radical expressions, absolute value equations and inequalities, sequences and patterns, systems of equations, quadratic inequalities, functions, modeling, matrices, roots of polynomials, and complex numbers.

[d]*Coordinate Geometry.* Items in this content area are based on graphing and the relations between equations and graphs, including points, lines, polynomials, circles, and other curves; graphing inequalities; slope; parallel and perpendicular lines; distance; midpoints; and conics.

[e]*Plane Geometry.* Items in this content area are based on the properties and relations of plane figures, including angles and relations among perpendicular and parallel lines; properties of circles, triangles, rectangles, parallelograms, and trapezoids; transformations; the concept of proof and proof techniques; volume; and applications of geometry to three dimensions.

[f]*Trigonometry.* Items in this content area are based on understanding trigonometric relations in right triangles; values and properties of trigonometric functions; graphing trigonometric functions; modeling using trigonometric functions; use of trigonometric identities; and solving trigonometric equations.

Table 2.19: Content Specifications for the ACT Reading Test

The items in the Reading Test are based on the prose passages that are representative of the kinds of writing commonly encountered in college freshman curricula, including prose fiction, the social sciences, the humanities, and the natural sciences. The four content areas and the approximate proportion of the test devoted to each are given below.

Reading passage content	Proportion of test	Number of items
Prose Fiction[a]	0.25	10
Social Science[b]	0.25	10
Humanities[c]	0.25	10
Natural Science[d]	0.25	10
Total	**1.00**	**40**

Scores reported: Social Studies/Sciences (Social Science, Natural Science)
Arts/Literature (Prose Fiction, Humanities)
Total test score

[a]*Prose Fiction*. The items in this category are based on short stories or excerpts from short stories or novels.

[b]*Social Science*. The items in this category are based on passages in the content areas of anthropology, archaeology, biography, business, economics, education, geography, history, political science, psychology, and sociology.

[c]*Humanities*. The items in this category are based on passages from memoirs and personal essays and in the content areas of architecture, art, dance, ethics, film, language, literary criticism, music, philosophy, radio, television, and theater.

[d]*Natural Science*. The items in this category are based on passages in the content areas of anatomy, astronomy, biology, botany, chemistry, ecology, geology, medicine, meteorology, microbiology, natural history, physiology, physics, technology, and zoology.

Table 2.20: Content Specifications for the ACT Science Test

The Science Test is based on the type of content that is typically covered in high school science courses. Materials are drawn from the biological sciences, the Earth/space sciences, physics, and chemistry. The test emphasizes scientific reasoning skills rather than recall of specific scientific content, skill in mathematics, or skill in reading. Minimal arithmetic and algebraic computations may be required to answer some items. The three formats and the approximate proportion of the test devoted to each are given below.

Content area[a]	Format	Proportion of test	Number of items
Biology Earth/Space Sciences Physics Chemistry	Data Representation[b]	0.38	15
	Research Summaries[c]	0.45	18
	Conflicting Viewpoints[d]	0.17	7
Total		**1.00**	**40**

Score reported: Total test score

[a]All four content areas are represented in the test. The content areas are distributed over the different formats in such a way that at least one passage, and no more than two passages, represents each content area.

[b]*Data Representation.* This format presents students with graphic and tabular material similar to that found in science journals and texts. The items associated with this format measure skills such as graph reading, interpretation of scatterplots, and interpretation of information presented in tables, diagrams, and figures.

[c]*Research Summaries.* This format provides students with descriptions of one or more related experiments. The items focus on the design of experiments and the interpretation of experimental results.

[d]*Conflicting Viewpoints.* This format presents students with expressions of several hypotheses or views that, being based on differing premises or on incomplete data, are inconsistent with one another. The items focus on the understanding, analysis, and comparison of alternative viewpoints or hypotheses.

Selection of Item Writers

Each year, ACT contracts with item writers to construct items for the ACT. The item writers are content specialists in the disciplines measured by the ACT tests. Most are actively engaged in teaching at various levels, from high school to university, and at a variety of institutions, from small private schools to large public institutions. ACT makes every attempt to include item writers who represent the diversity of the population of the United States with respect to ethnic background, gender, and geographic location.

Before being asked to write items for the ACT tests, potential item writers are required to submit a sample set of materials for review. Each item writer receives an item writer's guide that is specific to the content area. The guides include examples of items and provide item writers with the test specifications and ACT's requirements for content and style. Included are specifications for fair portrayal of all groups of individuals, avoidance of subject matter that may be unfamiliar to members of certain groups within society, and nonsexist use of language.

Each sample set submitted by a potential item writer is evaluated by ACT Test Development staff. A decision concerning whether to contract with the item writer is made on the basis of that evaluation.

Each item writer under contract is given an assignment to produce a small number of multiple-choice items. The small size of the assignment ensures production of a diversity of material and maintenance of the security of the testing program, since any item writer will know only a small proportion of the items produced. Item writers work closely with ACT test specialists, who assist them in producing items of high quality that meet the test specifications.

Item Construction

The item writers must create items that are educationally important and psychometrically sound. A large number of items must be constructed because, even with good writers, many items fail to meet ACT's standards.

Each item writer submits a set of items, called a *unit*, in a given content area. Most Mathematics Test items are discrete (not passage-based), but occasionally some may belong to sets composed of several items based on the same paragraph or chart. All items on the English and Reading Tests are related to prose passages. All items on the Science Test are related to passages and/or other stimulus material (such as graphs and tables).

Review of Items

After a unit is accepted, it is edited to meet ACT's specifications for content accuracy, word count, item classification, item format, and language. During the editing process, all test materials are reviewed for fair portrayal and balanced representation of groups within society and for nonsexist use of language. The unit is reviewed several times by ACT staff to ensure that it meets all of ACT's standards.

Copies of each unit are then submitted to content and fairness experts for external reviews prior to the pretest administration of these units. The content review panel consists of high school teachers, curriculum specialists, and college and university faculty members. The content panel reviews the unit for content accuracy, educational importance, and grade-level appropriateness. The fairness review panel consists of experts in diverse educational areas who represent both genders and a variety of racial and ethnic backgrounds. The fairness panel reviews the unit to help ensure fairness to all examinees. Any comments on the units by the content consultants are discussed in a panel meeting with all the content consultants and ACT staff, and appropriate changes are made to the unit(s). All fairness consultants' comments are reviewed and discussed, and appropriate changes are made to the unit(s).

Item Tryouts

The items that are judged to be acceptable in the review process are assembled into tryout units for pretesting on samples from the national examinee population. These samples are carefully selected to be representative of the total examinee population. Each sample is administered a tryout unit from one of the four academic areas covered by the ACT tests. The time limits for the tryout units permit the majority of students to respond to all items.

Item Analysis of Tryout Units

Item analyses are performed on the tryout units. For a given unit the sample is divided into low-, medium-, and high-performing groups by the individuals' scores on the ACT test in the same content area (taken at the same time as the tryout unit). The cutoff scores for the three groups are the 27th and the 73rd percentile points in the distribution of those scores. These percentile points maximize the critical ratio of the difference between the mean scores of the upper and lower groups, assuming that the standard error of measurement in each group is the same and that the scores for the entire examinee population are normally distributed (Millman & Greene, 1989).

Proportions of students in each of the groups correctly answering each tryout item are tabulated, as well as the proportion in each group selecting each of the incorrect options. Biserial and point-biserial correlation coefficients between each item score (correct/incorrect) and the total score on the corresponding test of the regular (national) test form are also computed.

Item analyses serve to identify statistically effective test items. Items that are either too difficult or too easy, and items that fail to discriminate between students of high and low educational achievement as measured by their corresponding ACT test scores, are eliminated or revised for future item tryouts. The biserial and point-biserial correlation coefficients, as well as the differences between proportions of students answering the item correctly in each of the three groups, are used as indices of the discriminating power of the tryout items.

Each item is reviewed following the item analysis. ACT staff members scrutinize items flagged for statistical reasons to identify possible problems. Some items are revised and placed in new tryout units following further review. The review process also provides feedback that helps decrease the incidence of poor quality items in the future.

Assembly of New Forms

Items that are judged acceptable in the review process are placed in an item pool. Preliminary forms of the ACT tests are constructed by selecting from this pool items that match the content and statistical specifications for the tests.

For each test in the battery, items for the new forms are selected to match the content distribution for the tests shown in Tables 2.17–2.20. Items are also selected to comply with the statistical specifications described on page 33. The distributions of item difficulty levels obtained on recent forms of the four tests are displayed in Table 2.21. The data in Table 2.21 are taken from random samples of approximately 2,000 students from each of the six national test dates during the 2010–2011 academic year. In addition to the item difficulty distributions, item discrimination indices in the form of observed mean biserial correlations and completion rates are reported.

Table 2.21: Difficulty[a] Distributions and Mean Discrimination[b] Indices for ACT Test Items, 2010–2011

	Observed difficulty distributions (frequencies)			
	English	Mathematics	Reading	Science
Difficulty range				
0.00–0.09	0	0	0	0
0.10–0.19	0	7	0	0
0.20–0.29	7	29	2	11
0.30–0.39	23	58	17	36
0.40–0.49	41	48	39	43
0.50–0.59	77	69	54	46
0.60–0.69	98	52	47	43
0.70–0.79	107	46	49	37
0.80–0.89	86	45	30	20
0.90–1.00	11	6	2	4
Number of items[c]	450	360	240	240
Mean difficulty	0.66	0.55	0.61	0.56
Mean discrimination	0.59	0.63	0.59	0.51
Avg. completion rate[d]	91	93	94	94

[a] Difficulty is the proportion of examinees correctly answering the item.

[b] Discrimination is the item-total score biserial correlation coefficient.

[c] Six forms consisting of the following number of items per test: English 75, Mathematics 60, Reading 40, Science 40.

[d] Mean proportion of examinees who answered each of the last five items.

The average completion rate is an indication of how speeded a test is for a group of students. A test is considered to be speeded if most students do not have sufficient time to answer the items in the time allotted. The completion rate reported in Table 2.21 for each test is the average completion rate for the six national test dates during the 2010–2011 academic year. The completion rate for each test is computed as the average proportion of examinees who answered each of the last five items.

Content and Fairness Review of Test Forms

The preliminary versions of the test forms are subjected to several reviews to ensure that the items are accurate and that the overall test forms are fair and conform to good test construction practice. The first review is performed by ACT staff. Items are checked for content accuracy and conformity to ACT style. The items are also reviewed to ensure that they are free of clues that could allow testwise students to answer the item correctly even though they lack knowledge in the subject areas or the required skills.

The preliminary versions of the test forms are then submitted to content and fairness experts for external review before the operational administration of the test forms. These experts are different individuals from those consulted for the content and fairness reviews of tryout units.

Two panels, a content review panel and a fairness review panel, are then convened to discuss with ACT staff the consultants' reviews of the forms. The content review panel consists of high school teachers, curriculum specialists, and college and university faculty members. The content panel reviews the forms for content accuracy, educational importance, and grade-level appropriateness. The fairness review panel consists of experts in diverse areas of education who represent both genders and a variety of racial and ethnic backgrounds. The fairness panel reviews the forms to help ensure fairness to all examinees.

After the panels complete their reviews, ACT summarizes the results. All comments from the consultants are reviewed by ACT staff members, and appropriate changes are made to the test forms. Whenever significant changes are made, the revised components are again reviewed by the appropriate consultants and by ACT staff. If no further corrections are needed, the test forms are prepared for printing.

In all, at least sixteen independent reviews are made of each test item before it appears on a national form of the ACT. The many reviews are performed to help ensure that each student's level of achievement is accurately and fairly evaluated.

Review Following Operational Administration

After each operational administration, item analysis results are reviewed for any anomalies such as substantial changes in item difficulty and discrimination indices between tryout and national administrations. Only after all anomalies have been thoroughly checked and the final scoring key approved are score reports produced. Examinees may challenge any items that they feel are questionable. Once a challenge to an item is raised and reported, the item is reviewed by content specialists in the content area assessed by the item. In the event that a problem is found with an item, actions are taken to eliminate or minimize the influence of the problem item as necessary. In all cases, the person who challenges an item is sent a letter indicating the results of the review.

Also, after each operational administration, DIF (differential item functioning) analysis procedures are conducted on the test data. DIF can be described as a statistical difference between the probability of the specific population group (the "focal" group) getting the item right and the comparison population group (the "base" group) getting the item right given that both groups have the same level of achievement with respect to the content being tested. The procedures currently used for the analysis include the standardized difference in proportion-correct (STD) procedure and the Mantel-Haenszel common odds-ratio (MH) procedure.

Both the STD and MH techniques are designed for use with multiple-choice items, and both require data from significant numbers of examinees to provide reliable results. For a description of these statistics and their performance overall in detecting DIF, see the ACT Research Report entitled *Performance of Three Conditional DIF Statistics in Detecting Differential Item Functioning on Simulated Tests* (Spray, 1989). In the analysis of items in an ACT form, large samples representing examinee groups of interest (e.g., males and females) are selected from the total number of examinees taking the test. The examinees' responses to each item on the test are analyzed using the STD and MH procedures. Compared with preestablished criteria, the items with STD or MH values exceeding the tolerance level are flagged. The flagged items are then further reviewed by the content specialists for possible explanations of the unusual STD or MH results. In the event that a problem is found with an item, actions will be taken as necessary to eliminate or minimize the influence of the problem item.

ACT Scoring Procedures

For each of the four multiple-choice tests in the ACT (English, Mathematics, Reading, and Science), the raw scores (number of correct responses) are converted to scale scores ranging from 1 to 36.

The Composite score is the average of the four scale scores rounded to the nearest whole number (fractions of 0.5 or greater round up). The minimum Composite score is 1; the maximum is 36.

In addition to the four ACT test scores and Composite score, seven subscores are reported: two each for the English Test and the Reading Test and three for the Mathematics Test. As is done for each of the four tests, the raw scores for the subscore items are converted to scale scores. These subscores are reported on a score scale ranging from 1 to 18. The four test scores and seven subscores are derived independently of one another. The subscores in a content area do not necessarily add to the test score in that area.

Electronic scanning devices are used to score the four multiple-choice tests of the ACT, thus minimizing the potential for scoring errors. If a student believes that a scoring error has been made, ACT hand-scores the answer document (for a fee) upon receipt of a written request from the student. A student may arrange to be present for hand-scoring by contacting one of ACT's regional offices, but must pay whatever extra costs may be incurred in providing this special service. Strict confidentiality of each student's record is maintained.

For certain test dates (specified in the current year's booklet *Registering for the ACT*), examinees may obtain (upon payment of an additional fee) a copy of the test items used in determining their scores, the correct answers, a list of their answers, and a table to convert raw scores to the reported scale scores. For an additional fee, a student may also obtain a copy of his or her answer document. These materials are available only to students who test during regular administrations of the ACT on specified national test dates. If for any reason ACT must replace the test form scheduled for use at a test center, this offer is withdrawn and the student's fee for this optional service is refunded.

ACT reserves the right to cancel test scores when there is reason to believe the scores are invalid. Cases of irregularities in the test administration process—falsifying one's identity, impersonating another examinee (surrogate testing), unusual similarities in answers of examinees at the same test center, or other indicators that the test scores may not accurately reflect the examinee's level of educational achievement, including but not limited to examinee misconduct—may result in ACT's canceling the test scores. When ACT plans to cancel an examinee's test scores, it always notifies the examinee prior to taking this action. This notification includes information about the options available regarding the planned score cancellation, including procedures for appealing this decision. In all instances, the final and exclusive remedy available to examinees who want to appeal or otherwise challenge a decision by ACT to cancel their test scores is binding arbitration through written submissions to the American Arbitration Association. The issue for arbitration shall be whether ACT acted reasonably and in good faith in deciding to cancel the scores.

Technical Characteristics of the ACT Tests

The technical characteristics—the score scale, norms, equating, reliability, and validity—of the ACT Test is thoroughly documented in the *ACT Technical Manual* (ACT, 2007). The *ACT Technical Manual* can be acquired from ACT's website at **www.act.org**.

Chapter 3
Evidence of the Use of Procedures for Sensitivity and Bias Reviews and DIF Analyses

Commitment to Fairness

The purposes of this chapter are to 1) describe the sensitivity and bias procedures followed during development of the PSAE components that ensure that the tests are as fair as possible to all examinees who take them, and 2) to describe the analyses routinely executed after each operational administration that provide empirical evidence that the PSAE tests operated in a fair and unbiased manner.

The critical goal is to accurately assess what students can do with what they know in the content areas covered by the PSAE tests. If factors other than the academic skills and knowledge in those content areas were allowed to intrude, we would provide a less accurate picture of what students know and can do and would risk subjecting students to situations in which their performance might be adversely affected by language or contexts that are perceived to be unfair. ISBE is deeply committed to fairness in principle and in the interest of accuracy of the PSAE.

The *Code of Fair Testing Practices in Education is* a set of guidelines for those who develop, administer, and use educational tests and data, sets forth criteria for fairness in four areas: developing and selecting appropriate tests, administering and scoring tests, reporting and interpreting test results, and informing test takers. According to the *Code*, test developers should provide "tests that are fair to all test takers regardless of age, gender, disability, race, ethnicity, national origin, religion, sexual orientation, linguistic background, or other personal characteristics." Test developers should "avoid potentially insensitive content or language," and "evaluate the evidence to ensure that differences in performance are related to the skills being assessed." Development of the PSAE follows these standards for appropriate test development practice and use.

PSAE development also follows the *Code of Professional Responsibilities in Educational Measurement*, which numbers among test developers' responsibilities "to develop assessment products and services that are as free as possible from bias due to characteristics irrelevant to the construct being measured, such as gender, ethnicity, race, socioeconomic status, disability, religion, age, or national origin." To ensure fairness in a test is a critically important goal. Unfairness must be detected, eliminated, and prevented at all stages of test development, test administration, and test scoring. The work of ensuring test fairness starts with the design of the test and test specifications. It then continues through every stage of the test development process, including item (test question) writing and review, item pre-testing, item selection and forms construction, and forms review. Every effort is made to see that PSAE tests are fair for all Illinois students.

Fairness and Bias Reviews

To ensure fairness for all examinees, fairness concerns are systematically and continuously addressed throughout every stage of the test development process, from initial item writer recruitment, continuing throughout all steps until final PSAE tests are produced. By building fairness into all steps of the test development process, any concerns can be addressed immediately, thus significantly reducing risks of any fairness problems in the final test materials.

Fairness is a top consideration when recruiting and considering item writers. When selecting item writers, their demographic data and the demographic data of students they teach must be representative of Illinois's diverse student population. To ensure item writers write fair and unbiased items, Item Writer's Guides are immediately sent to item writers that explain in great detail how to write accurate and fair test material. Item writers are to assure that all test material they develop is judged to be appropriate for and equally familiar or unfamiliar to examinees of both sexes, and all geographic, socioeconomic, racial, ethnic, and cultural backgrounds. No examinee group should be placed at an advantage or disadvantage due to experience (or lack thereof) with a topic that is not central to the content or skill being measured. Item writers' submissions that do not meet any of these criteria will be rejected.

Upon acceptance of item writers' submissions, all PSAE test materials are subjected to several quality control and sensitivity reviews to ensure that the test materials are fair and conform to good test construction practice. Test materials are submitted to fairness experts for external review before the operational administration of the test forms. Fairness and bias experts carefully review each item and prompt to ensure that neither the language nor the content of the test material will be offensive to a test taker, and that no item will disadvantage any student from any geographic, socioeconomic, or cultural background.

After the consultants complete their reviews, comments from the consultants are reviewed by PSAE test developers and appropriate changes are made to the test material. Whenever significant changes are made, the revised components are again reviewed by the appropriate consultants and by PSAE test developers. In all, multiple independent reviews are made of each test item before it appears on a PSAE test form. Several different independent reviews are performed of each PSAE component to help ensure that each student's level of achievement is accurately and fairly evaluated.

Differential Item Functioning Analysis

To check for item bias, multiple-choice tryout items and operational items are analyzed for differential item functioning (DIF). DIF can be described as a statistical difference between the probability of a specific population group (the "focal" group) getting the item right and a comparison population group (the "base" group) getting the item right given that both groups have the same level of achievement with respect to the content being tested. Following any PSAE administration, DIF analyses are performed on all items.

The procedures currently used for DIF analyses include the Mantel-Haenszel common odds-ratio (MH) procedure and the standardized difference in proportion-correct (STD) procedure. Both the MH and STD techniques are designed for use with multiple-choice items, and both require data from significant numbers of examinees to provide reliable results. For a description of these statistics and their performance overall in detecting DIF, see the ACT Research Report entitled *Performance of Three Conditional DIF Statistics in Detecting Differential Item Functioning on Simulated Tests* (Spray, 1989).

In the analysis of items, large samples representing focal and base groups of interest (e.g., females and males) are selected from the total number of examinees taking the test. The examinees' responses to each operational ACT item and WorkKeys item are analyzed using both the MH and STD procedures. Items with MH alpha or STD values exceeding pre-established tolerance levels (i.e., MH alpha values less than or equal to 0.5, MH alpha values greater than or equal to 2.0, or STD values greater than or equal to 0.1 in absolute value) are flagged for review.

Responses to ISBE-developed science test operational and tryout items are analyzed using the MH delta statistic at a significance level of 0.05. Each ISBE-developed science test item is classified into one of three categories: A (negligible DIF), B (moderate DIF), and C (large DIF). An item is classified in category A if the MH delta value is not statistically different from zero or if the MH delta value is less than 1.0 in absolute value. An item is classified in category C if the MH delta value is statistically different from zero and is greater than 1.5 in absolute value. All other items are classified in category B. All category C items are flagged for review.

All flagged ACT, WorkKeys, and ISBE-developed science test items are reviewed by PSAE test developers for possible explanations for the unusual results. In the event that a problem is found with an item, actions are taken as necessary to eliminate or minimize the influence of the problem item. Flagged tryout items that are judged to be problematic are not used in subsequent test form construction. It should be noted that the act of flagging an item does not mean the item is necessarily unfair.

Regarding analytical techniques employed on writing prompts, once scoring of the Writing Test prompts has been completed, the prompts are analyzed for acceptability, validity, and accessibility. The prompts are also reviewed to ensure that they are compatible with previous operational prompts and that they function in the same way as previous prompts.

A summary of the DIF analysis results for the PSAE Standard form administered in 2012 is shown in Table 3.1, which provides the number of comparisons by group favored that were flagged by (1) Either MH or STD or both (for ACT and WorkKeys only) or by (2) "C"-Level DIF (for ISBE-developed science only).

Table 3.1: Summary of DIF Analysis Results for the PSAE Standard Form Administered in Spring 2012

	Subject		
Favored group	**Reading**	**Mathematics**	**Science**
Male		1	1
Female			
African American		1	1
Caucasian			
Hispanic American			
Caucasian			

Table 3.1 indicates that in Mathematics, for example, 1 out of the 90 items administered on the standard form appeared to favor males while 1 item appears to favor African Americans, based on the statistical indices. A total of 4 out of the 720 comparisons made on all PSAE standard form items were flagged and further reviewed by content and measurement specialists. The reviewers concluded that no gender, cultural, or racial bias was evident in the test items and that the item content was consistent with Illinois Learning Standards.

Chapter 4
Scaling, Reliability, and Measurement Error of the PSAE

PSAE scale scores are reported for reading, mathematics, and science. All three of these scales are based on combinations of two assessments. The following descriptions pertain to the PSAE reading, mathematics, and science scales.

The range of scores on the PSAE scales is 120 to 200 with an increment of 1. The target means and standard deviations of the PSAE score scale were 160 and 15, respectively, for each of the three scores. The means and standard deviations pertain to grade 11 students in Illinois public schools.

Scaling of the PSAE Reading, Mathematics, and Science Assessments

Over 110,000 grade 11 students in Illinois public schools took the PSAE assessment in April and May 2001. A selected sample of 10,554 students who took the PSAE assessment in April, referred to in this report as the "scaling group," was used in creating the PSAE reading, mathematics, and science scales. This section contains a discussion of the data used in scaling the PSAE.

The Scaling Process

Based on feedback from peer reviewers to obtain increased alignment between the PSAE and the Illinois Learning Standards, it was decided to compute PSAE scores directly from item scores rather than weighting component scores, as was done in the previous scaling study. It was suggested that an IRT approach be used to maintain PSAE scores, instead of classical methodology. The IRT methodology was initiated on Mathematics, Reading, and Science in spring 2008.

To ensure the PSAE scores obtained from the new methodology are interchangeable with those from the original methodology, a bridge study was conducted to link scores from both methodologies. The impact of the new methodology was examined in the same study.

The 2007 initial form data were chosen for the bridge study. For each examinee, the PSAE raw score was computed by summing up the raw scores of the two components (Day 1 and Day 2). In order to have the same percentage of students at each score point using the original and new scoring methods, equipercentile concordance was conducted between these PSAE raw scores and PSAE original scale scores resulting in a raw-to-scale score conversion table.

The raw-to-scale-score transformations of the PSAE assessment components obtained in the bridge study and used as the basis for the 2008 scaling are presented in Figures 4.1–4.3. The raw-to-scale-score transformations are approximately linear in the middle part of the scale score ranges for the PSAE Reading and Science scales and approximately arcsine for Mathematics. The transformations are flat at extremely low scores because of truncations. At extremely high scores, the transformation for Mathematics is also truncated to the highest possible score, 200. These findings are consistent with those in the 2001 scaling study.

Figure 4.1: Raw-to-Scale-Score Transformation for PSAE Reading

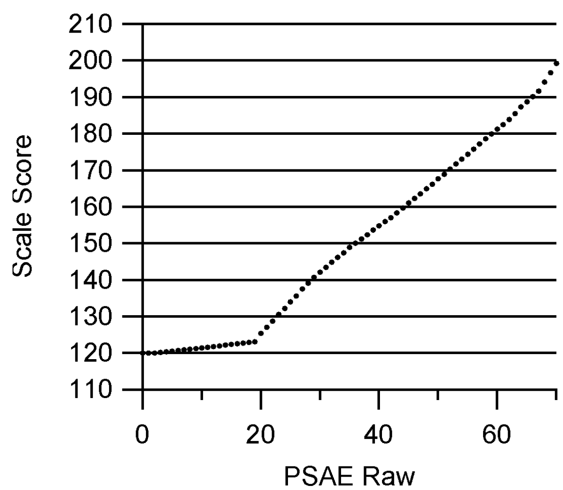


Figure 4.2: Raw-to-Scale-Score Transformation for PSAE Mathematics

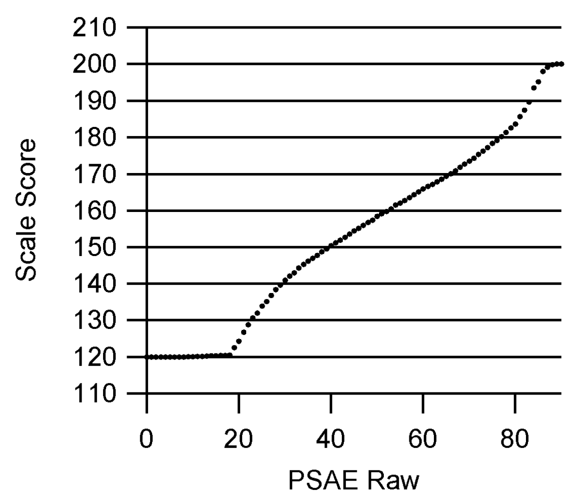
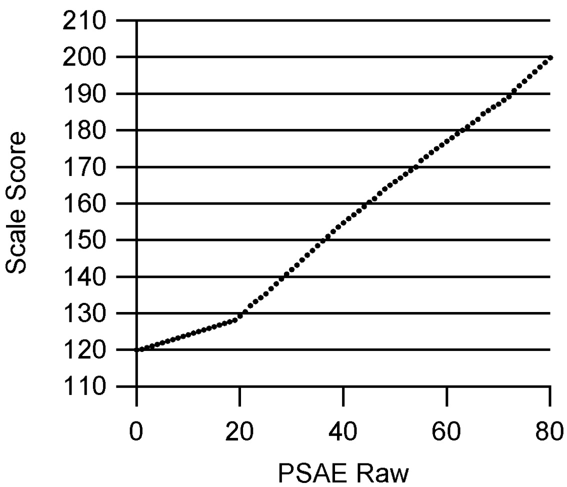


Figure 4.3: Raw-to-Scale-Score Transformation for PSAE Science



Summary Statistics

Scale-score summary statistics for the bridge study group are provided in Table 4.1 for the PSAE scale scores. The scale-score means and standard deviations of the PSAE scales were close to those from the 2001 scaling study, which were reported in the 2007 PSAE Technical Manual (ISBE, 2007).

Table 4.1: Scale-Score Summary Statistics for the PSAE Scales for the Bridge Study Group

Statistics	Reading	Mathematics	Science
Mean	158.5085	159.1001	159.7703
SD	14.8818	15.6125	14.2794
Skewness	0.0824	0.2079	–0.0290
Kurtosis	–0.5129	–0.0507	–0.6647
N	114,882	114,902	114,546

Linking

PSAE Reading, Mathematics, and Science are each made up of two separately timed component tests. Of these six component tests, one has common items across different forms, two may or may not have common items across forms, and three do not have common items across forms. Therefore, the linking across PSAE forms cannot rely only on common item equating. Using non-PSAE data, different forms of the ACT tests can be put on a common scale using a random groups design and IRT methodology.

The ACT items in PSAE Forms 1 (initial form 2007) and 2 (say, initial form 2008) can be placed on the common PSAE IRT scale by using the non-PSAE ACT equating data (i.e., all ACT items can be placed on a common scale, which can then be scaled to the PSAE scale for PSAE Form 1, thus resulting in all ACT item IRT parameter estimates being scaled to the PSAE IRT scale). A commonly used method in the industry, the Stocking-Lord method (Stocking & Lord, 1983), was used to place all ACT items on a common scale and on the PSAE scale. As directed by ISBE, the ACT item pool was used as a bridge to link between 2008 forms and 2007 forms. For example, for PSAE Reading, all 40 ACT Reading items and 30 WorkKeys Reading items were calibrated together in a single run. The Stocking-Lord constants were found by comparing the ACT item parameter estimates from this run to the previously scaled values. Using these constants, all 80 PSAE Reading items were placed on the PSAE IRT scale.

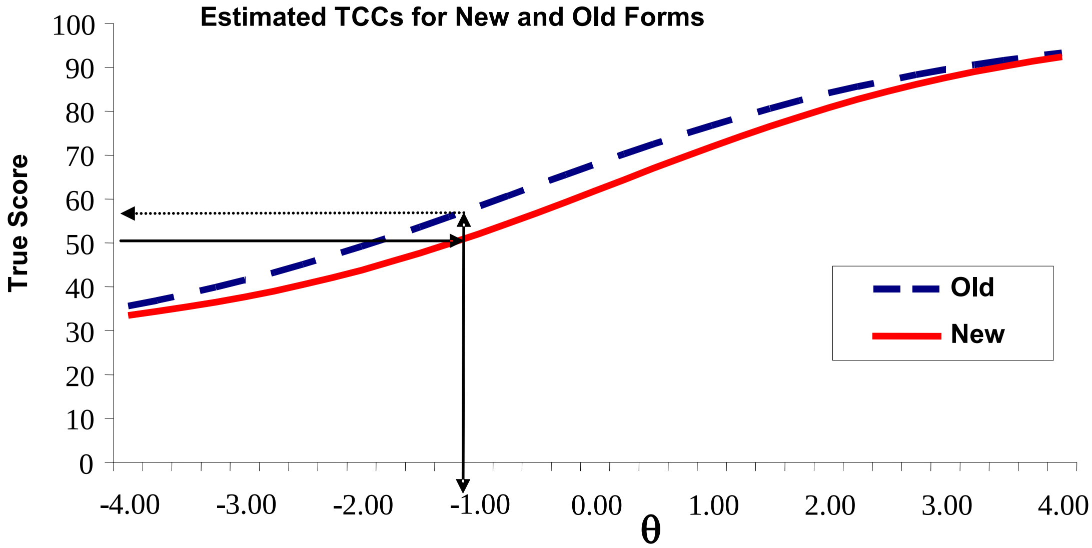
IRT Equating

The rescaled item parameter values were used in an IRT true score equating procedure (Kolen & Brennan, 2004) to equate raw scores on 2008 forms to raw scores on 2007 forms. In this procedure, the rescaled item parameters were used to produce test characteristic curves (TCCs) and the true score associated with a given theta on a 2008 form (new form) was considered to be equivalent to the true score associated with that theta on a 2007 form (old form). Figure 4.4 shows how to find the equated score on the old form of a true score of 50 on the new form. Using the TCC for the new form, we can find the theta value of –1.00 is associated with a true score of 50 on the new form. Using the TCC for the old form, we can find a true score of 57.2 is associated with the same theta value of –1.00. Because they are associated with the same theta value, 57.2 is the equated raw score on the old form of a true score of 50 on the new form.

Creating Raw-to-Scale Conversion Tables

Because the equated raw scores on a 2008 form are interchangeable with the raw scores on a 2007 form, the equated raw scores were used to look up the PSAE scale scores in the 2007 raw-to-scale conversion tables to create the 2008 raw-to-scale conversion tables. Since the equated raw scores are typically not integer whereas the raw scores in the 2007 raw-to-scale conversion tables are integer, we used the linear interpolation method to find the PSAE scale score corresponding to a non-integer raw score. Consistent with what has been done previously, the top PSAE raw scores were converted to the top PSAE scale score, 200.

Figure 4.4: An Example of IRT True Score Equating



2012 Item Calibration

The data for the calibration were obtained from combining both Day 1 and Day 2 data. All students who met attemptedness for PSAE were included in the PSAE calibrations. The included students had to take the same type of administration forms for both Day 1 and Day 2 (i.e., if the Day 1 administration form is an initial form, the Day 2 administration form has also to be an initial form). The reason for the requirement of the same type of administration forms is that the sample sizes for other combinations (e.g., Day 1 initial plus Day 2 makeup) were too small to be calibrated appropriately. Calibration started when it was determined that (a) a sufficient sample size was available given the number of students who were administered a form and/or (b) waiting for additional examinees would jeopardize the schedule.

Table 4.2 summarizes the results of the calibration of the 2012 data. As shown in this table, all calibrations converged in a range of 20 through 65 cycles. Table 4.2 also shows that the number of misfit items ranges from 1 to 86. It is obvious that the number of misfit items correlates with the sample size (e.g., the initial form has the largest sample size as well as the largest number of misfit items; while the makeup form has the smallest sample size as well as the smallest number of misfit items). This finding is not surprising because it is well known that the chi-square statistic as used in BILOG-MG is sensitive to the sample size.

Table 4.2: Convergence and Item Fit

				Item fit			
				P < .05		P < .01	
Form	Test	Number of calibration cycles	Total number of items	N of good fit	N of misfit	N of good fit	N of misfit
Initial	Mathematics	24	90	4	86	5	85
	Reading	20	70	2	68	6	64
	Science	22	80	7	73	11	69
Makeup	Mathematics	37	90	85	5	89	1
	Reading	27	70	69	1	69	1
	Science	38	80	69	11	77	3
Accommodated	Mathematics	65	90	22	68	32	58
	Reading	36	70	27	43	38	32
	Science	46	80	7	73	15	65

Measurement Error and Reliability for the PSAE Scores

The conditional standard errors of measurement (CSEM) summarize the amount of error or inconsistency of reported scores at different points on the score scale. Because the components of the PSAE Mathematics, Reading, and Science assessments contain only dichotomously scored items and these items are calibrated using an IRT model, the CSEM for raw scores are computed under the IRT framework (Lord, 1980). Given the CSEM for raw scores, the CSEM for PSAE scale scores are obtained through the delta method (Kendall & Stuart, 1977). In order for this method to work, polynomial models were fitted to the raw to scale conversion tables.

The estimated scale-score reliability for the assessment i, denoted (rel_i), where i = the PSAE Mathematics, Reading, or Science asessment, is calculated as

$$rel_i = 1 - \frac{\bar{\sigma}^2(E_i)}{\sigma^2(S_i)},$$

where $\bar{\sigma}^2(E_i)$ is the average of the estimated scale score conditional error variances and $\sigma^2(S_i)$ is the observed scale-score variance for test i. The mean, variance, average standard error of measurement, and reliability estimates for the PSAE Spring 2012 administration of the initial form are shown in Table 4.3. The CSEM for the PSAE scale scores are shown in Figures 4.5–4.7. The error and reliability statistics and CSEM plots look reasonable given the scale.

In 2012, fitting the polynomial used to approximate the raw-to-scale-score conversion was enhanced by excluding extremely low scores where the conversion is constant, and based on very little data. For example, in math, raw scores of 0 to 19 all converted to a scale score of 120. Hence, the polynomial approximation of the raw-to-scale-score conversion did not incorporate scores below 19 because there is no variability of the conversion in this range. This improved the approximation for students with scale scores above 120 on math.

Table 4.3: Average Standard Errors of Measurement (SEMs) and Reliabilities for the PSAE Spring 2012 Administration (Initial Form)

Statistics	Reading	Mathematics	Science
Scale score mean	156.81	158.67	159.85
Scale score variance	221.00	228.75	214.19
Average error variance	16.33	16.75	15.84
Scale Score SEM	4.04	4.09	3.98
Scale Score Reliability	0.93	0.93	0.93
N	125,718	125,736	125,731

Figure 4.5: PSAE Reading—Conditional Standard Errors of Measurement (CSEM) by Observed Scale Score for the PSAE Spring 2012 Administration

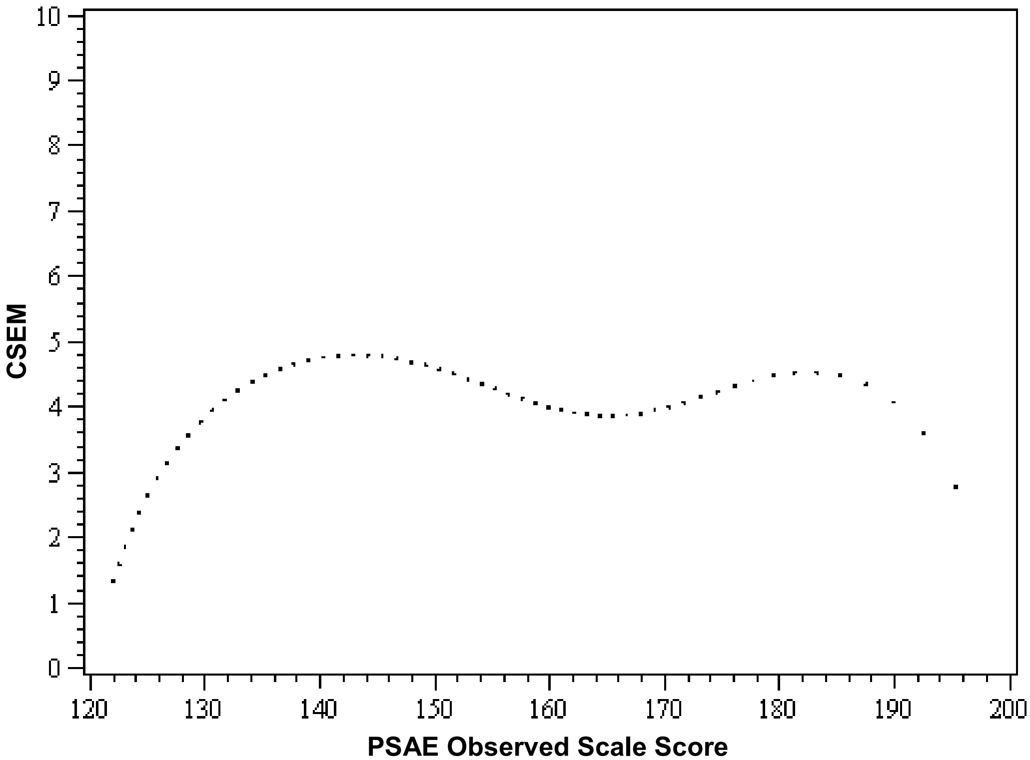


Figure 4.6: PSAE Mathematics—Conditional Standard Errors of Measurement (CSEM) by Observed Scale Score for the PSAE Spring 2012 Administration

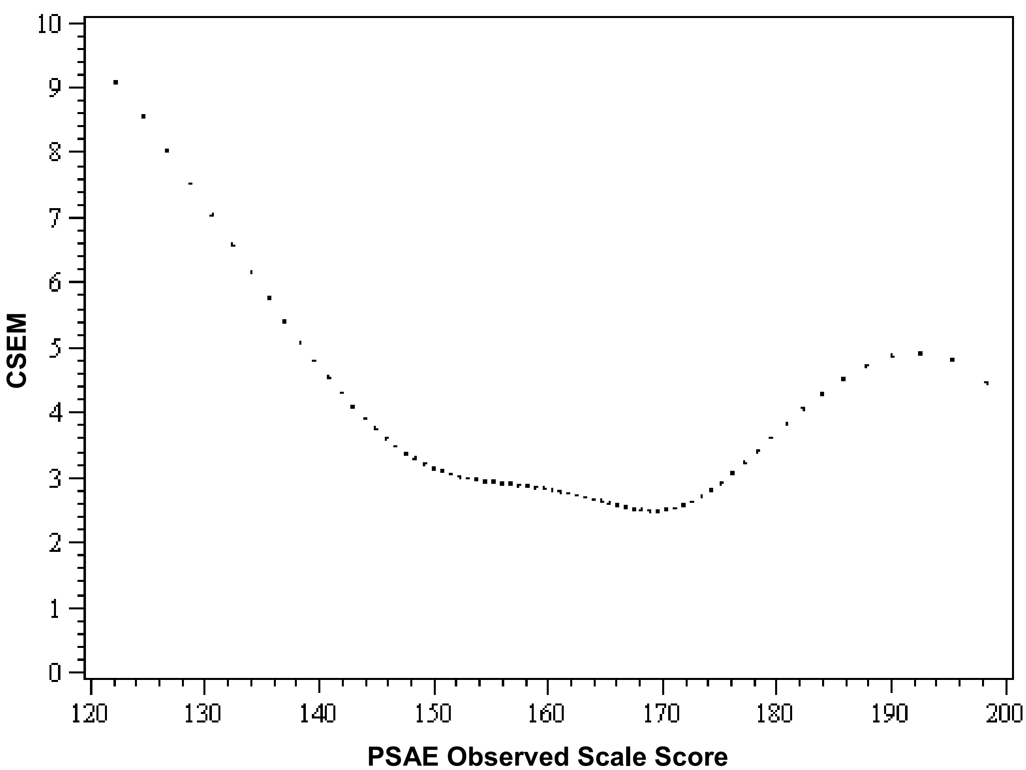
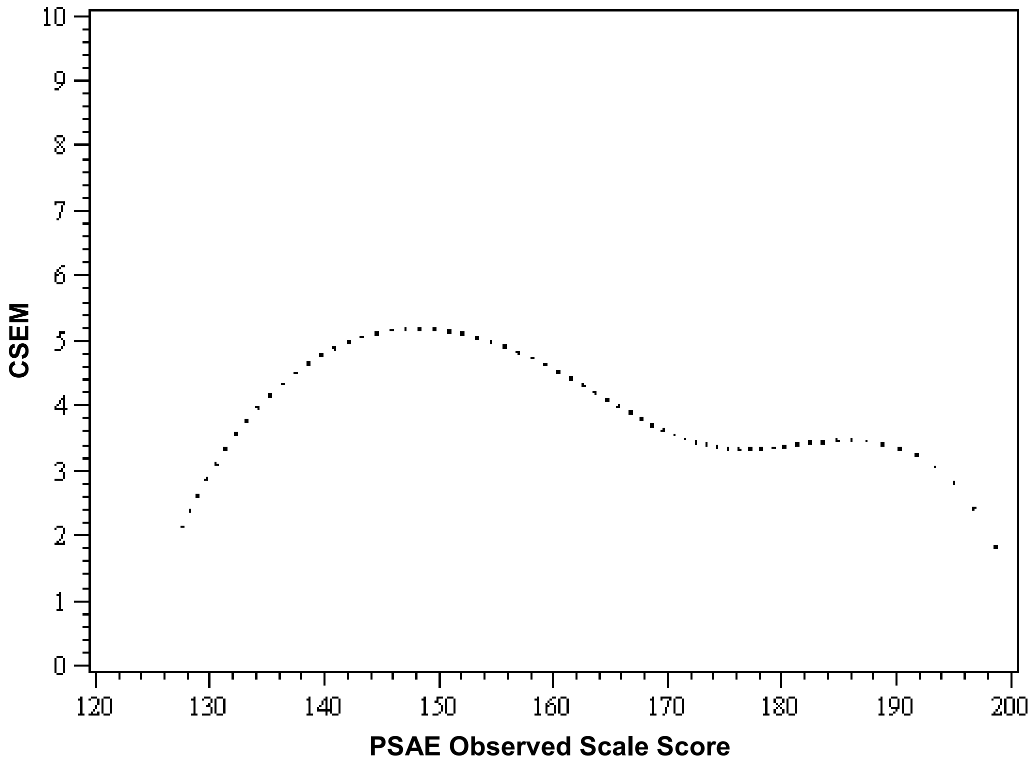


Figure 4.7: PSAE Science—Conditional Standard Errors of Measurement (CSEM) by Observed Scale Score for the PSAE Spring 2012 Administration



Chapter 5
Classification Consistency for the PSAE

Setting Standards on the PSAE

When administered for the first time in spring 2001, the PSAE assessed reading, mathematics, science, writing, and social science. In 2001, for each PSAE test, three cutoff score points and four categories at the scale-score level were established: Academic Warning, Below Standards, Meets Standards, and Exceeds Standards. A description of the 2001 standard-setting process in these subject areas can be found in Chapter 4 of each *Prairie State Achievement Examination Technical Manual* issued for 2001–2005 (ISBE, 2001, 2002, 2003, 2004, 2005). Due to changes in state law, writing and social science were no longer assessed beginning in 2005, and writing was assessed once again starting in 2007, but with a different PSAE assessment than was given in 2001–2004. The PSAE Writing Test administered in 2007 included the same multiple-choice component (the ACT English Test) as in previous years, but the ISBE-developed writing prompt was replaced by the ACT Writing Assessment. As a result, a new standard-setting process took place in 2007 for PSAE Writing in order to establish performance-level cutoff points based on this new assessment. A description of the standard-setting for the PSAE Writing Test can be found in Chapter 5 of the 2007 *Prairie State Achievement Examination Technical Manual* (ISBE, 2007). Changes in state law eliminated writing in 2012. Table 5.1 presents the PSAE scale score cut points in subject areas tested in 2012, as determined by the 2001 standard-settings.

Table 5.1: PSAE Scale Score Cut Points for Reading, Mathematics, and Science

Subject	Academic Warning (Level 1)	Below Standards (Level 2)	Meets Standards (Level 3)	Exceeds Standards (Level 4)
Reading	120–134	135–154	155–177	178–200
Mathematics	120–135	136–155	156–178	179–200
Science	120–135	136–157	158–177	178–200

2012 Classification Consistency

It has been typical to estimate classification consistency with a single test administration using a psychometric model (Hanson & Brennan, 1990; Livingston & Lewis, 1995) because the test (or parallel forms of the test) is not often administered twice to the same sample. As stated above, for each PSAE test, there are three cutoff score points and four categories at the scale-score level: Academic Warning, Below Standards, Meets Standards, and Exceeds Standards. Examinees are classified into one of the four mutually exclusive categories based on their scale scores and the cutoff points on the PSAE assessment. To estimate classification consistency, however, 4 × 4 contingency tables for the PSAE assessment are created using the psychometric model, with the columns and rows showing the four classification categories. The elements of the 4 × 4 tables indicate the joint probabilities of examinees being classified in the pairs of the column and row categories; for example, being classified in the Below Standards level on one occasion (column) and in the Meets Standards level on the other (row). The sums of the diagonal elements of the 4 × 4 tables are the indices of classification consistency.

The data used to compute classification consistency are based on examinees who took the initial form PSAE tests. An IRT procedure described by Lee (2010) was followed to compute classification consistency indices for Mathematics, Reading, and Science.

With this procedure, the distribution of abilities was estimated from the data and the expected conditional distributions of raw scores were computed given item parameter values. Accordingly, the probabilities of examinees being classified into each category were computed. Assuming a test-retest model with independent errors of measurement, the probabilities of being classified into each pair of categories (4 × 4) were computed. By summing the probabilities in the diagonal elements in the 4 × 4 tables, classification consistencies were estimated.

Tables 5.2–5.4 show the 4 × 4 contingency tables and indices of classification consistency for the PSAE assessments. The classification consistency indices vary over the PSAE assessments because of different measurement errors.

Table 5.2: Spring 2012 Classification Consistency for PSAE Reading

(N = 123,239)	Academic Warning	Below	Meets	Exceeds
Academic Warning	3%	2%	0%	0%
Below	2%	31%	6%	0%
Meets	0%	6%	36%	3%
Exceeds	0%	0%	3%	8%
		Classification Consistency: 78%		

Table 5.3: Spring 2012 Classification Consistency for PSAE Mathematics

(N = 123,252)	Academic Warning	Below	Meets	Exceeds
Academic Warning	3%	3%	0%	0%
Below	3%	30%	4%	0%
Meets	0%	4%	40%	2%
Exceeds	0%	0%	2%	9%
		Classification Consistency: 82%		

Table 5.4: Spring 2012 Classification Consistency for PSAE Science

(N = 123,256)	Academic Warning	Below	Meets	Exceeds
Academic Warning	2%	3%	0%	0%
Below	3%	30%	6%	0%
Meets	0%	6%	35%	3%
Exceeds	0%	0%	3%	9%
		Classification Consistency: 77%		

Chapter 6
Ensuring Consistency of PSAE Score Meaning Over Time

The PSAE program is administered in April, with a makeup administration in May. So that scores from these different administrations are comparable, as well as to allow tracking of trends across time, new forms of the PSAE must be related to older forms. The ACT, WorkKeys assessments, and the ISBE-developed science test must be placed on the PSAE score scales. This is accomplished by equating new forms of the tests to a form already on the underlying raw score scale.

To maintain PSAE scores over time, new forms of the components are developed to rigid, consistent content and statistical specifications, and the raw component scores for new forms are equated to the raw scores of the base form. These non-integer scores are then inserted into the raw-to-PSAE score conversions developed in the scaling study, which allows PSAE scores from 2012 to be compared to PSAE scores from prior years.

Equating of the ISBE-Developed Science Test

New forms of the ISBE-developed science test are equated using a common item design. In a common-item design, the new form has a set of items in common with a previously administered (and equated) form. The common items are chosen to represent the content and statistical characteristics of the test and are interspersed among the new items on the new form. The common items have estimated Rasch parameters that are on the "ISBE-developed science scale," due to their having appeared on the previously administered form, and having been calibrated and scaled at that time. When the data on the new form is calibrated, the common item parameters are fixed at their scaled values from the previous administration, and thus the common items serve to anchor the scaling of all the items on the new form.

Equating of WorkKeys Forms

New forms of the WorkKeys tests are developed to adhere to the same content and statistical specifications, however, the forms may be slightly different in difficulty. To control for these differences, scores on all forms are equated so that when they are reported to examinees, equated scale scores have the same meaning regardless of the particular form administered.

Two common equating designs that are used with the WorkKeys tests are the randomly equivalent groups design and the common-item nonequivalent groups design. In a *randomly equivalent groups design*, new test forms are administered along with an anchor form that has already been equated to previous forms. A spiraling process is used to distribute test forms to examinees. Thus, in each testing room the first person receives Form 1, the next Form 2, and the next Form 3. This pattern is repeated so that each form is given to one-third of the examinees and the forms are given to randomly equivalent groups. When this design is used, the difference in total-group performance on the new and anchor forms is considered a direct indication of the difference in difficulty between the forms. Scores on the new forms are equated using various equating methodologies including linear and equipercentile procedures.

The randomly equivalent groups design is commonly used for equating WorkKeys test forms. However, a *common-item nonequivalent groups design* has been used when a spiraling technique cannot be implemented in a test administration or when only a single form can be administered per test date. In a common-item nonequivalent groups design, the new form(s) and base form have a set of items in common, and different groups of examinees are administered the different forms. The common (anchor) item sets are chosen to represent the content and statistical characteristics of the test and are usually interspersed among the other items in the new test form.

In this design, the groups are not assumed to be equivalent. The common items are used to adjust for group differences. Observed differences between group performances can result from a combination of examinee group differences and test form differences. Strong statistical assumptions are usually required to separate these differences.

Equating of ACT Forms

Several new forms of the ACT are developed each year. Even though each form is constructed to adhere to the same content and statistical specifications, the forms may differ slightly in difficulty. To control for these differences, subsequent forms are equated, and the scores reported to examinees are scale scores that have the same meaning regardless of the particular form administered to examinees. Thus, scale scores are comparable across test forms and test dates.

A carefully selected sample of examinees from one of the five national test dates each year is used as an equating sample. The examinees in this sample are administered a spiraled set of "n" forms—the new forms ("n – 1" of them) and one anchor form that has already been equated to previous forms. (The anchor form is the form used initially to establish the score scale.) The use of randomly equivalent groups is an important feature of the equating procedure and provides a basis for confidence in the continuity of scales. More than 2,000 examinees take each form.

Scores on the alternate forms are equated to the score scale using equipercentile equating methodology. In equipercentile equating, a score on Form X of a test and a score on Form Y are considered to be equivalent if they have the same percentile rank in a given group of examinees. The equipercentile equating results are subsequently smoothed using an analytic method described by Kolen (1984) to establish a smooth curve, and the equivalents are rounded to integers. The conversion tables that result from this process are used to transform raw scores on the new forms to scale scores.

The equipercentile equating technique is applied to the raw scores of each of the four multiple-choice tests for each form separately. The Composite score is not directly equated across forms. It is, instead, a rounded arithmetic average of the scale scores for the four equated tests. The subscores are also separately equated using the equipercentile method. Note, in particular, that the equating procedure does not lead to a reported score for a test being equal to some prespecified arithmetic combination of subscores within that test.

As specified in the Standards for Educational and Psychological Testing (AERA, APA, NCME, 1999), ACT conducts periodic checks on the stability of the ACT scores. The results appear reasonably stable to date.

Comparing PSAE Scores Over Time

The equating of the separate components (ISBE Science, WorkKeys, and ACT) provides information on how the comparability of the scores contributing to the PSAE score are maintained over time. However, an external measure of the stability of PSAE would be useful to confirm this consistency. Future studies could make use of high school grades, college grades, and other variables external to the PSAE program. However, for an immediate check that requires no external variables, PSAE scores can be compared to scale scores on ISBE Science, WorkKeys, and ACT.

This analysis is admittedly somewhat confounded, as, for example, ISBE Science is a component of PSAE Science. However, PSAE Science scores are dependent on ISBE Science and ACT Science raw scores, not scale scores, and the scale scores have a long history of being stable over time. (For example, the scale for the ACT was last changed in 1989, when the test specifications were revised.)

For students who earned valid PSAE scores, Tables 6.1–6.6 provide information relating PSAE scores in reading, mathematics, and science to the component scale scores. The first column presents a component score (i.e., an ACT scale score, a WorkKeys level score, or an ISBE Science scale score), and the second column shows the approximate middle 90% of the distribution of PSAE scores associated with that component score. For example, in Table 6.1, 90% of the students who earned an ACT reading score of 21 received a PSAE reading score between 152 and 166. For students with a given component score, much of this variability in PSAE reading scores may be attributed to performance on the other component. Note that intervals containing fewer than 50 students would not be stable and are not reported. Columns 3, 4, and 5 in the tables compare the conditional mean PSAE scores over time in reading, mathematics, and science for 2012 and 2001. Column 5 presents the differences between the two sets of means. For example, in Table 6.1, an ACT score of 30 is associated with a PSAE score of 177 in 2012, and a score of 181 in 2001, a difference of four PSAE score points. Differences are small through the middle and upper ranges of the score scale but are a bit larger in the lower ranges of the scale, and this is true for the rest of the tables. This indicates that the scale is more stable where there are more examinees.

Table 6.1: Conditional Average PSAE Reading Means, Given Students' ACT Reading Scale Scores

ACT Reading	PSAE Reading 90% Interval	PSAE Reading 2012	PSAE Reading 2001	Difference (2012 – 2001)
1	—	122	121	1
2	—	123	130	–7
3	—	122	128	–6
4	—	124	120	4
5	—	122	127	–5
6	121–132	125	128	–3
7	121–133	125	129	–4
8	122–133	126	127	–1
9	122–135	127	130	–3
10	122–138	130	133	–3
11	122–140	131	136	–5
12	125–143	135	139	–4
13	130–147	139	142	–3
14	131–148	141	146	–5
15	137–151	145	149	–4
16	141–154	147	150	–3
17	143–156	150	153	–3
18	146–158	152	155	–3
19	148–161	154	157	–3
20	150–163	157	159	–2
21	152–166	159	162	–3
22	156–168	162	164	–2
23	158–170	164	166	–2
24	161–173	167	167	0
25	163–175	170	170	0
26	164–177	172	173	–1
27	167–179	174	174	0
28	170–181	175	177	–2
29	165–182	176	179	–3
30	171–183	177	181	–4
31	174–184	180	182	–2
32	175–187	182	183	–1
33	178–191	184	184	0
34	181–196	188	186	2
35	—	189	188	1
36	183–200	192	190	2

Table 6.2: Conditional Average PSAE Reading Means, Given Students' WorkKeys *Reading for Information* Level Scores

WK Reading	PSAE Reading 90% Interval	PSAE Reading 2012	PSAE Reading 2001	Difference (2012 – 2001)
0	122–137	127	125	2
3	125–144	133	133	0
4	132–157	144	147	–3
5	144–171	157	161	–4
6	156–183	170	174	–4
7	168–196	182	185	–3

Table 6.3: Conditional Average PSAE Mathematics Means, Given Students' ACT Mathematics Scale Scores

ACT Mathematics	PSAE Mathematics 90% Interval	PSAE Mathematics 2012	PSAE Mathematics 2001	Difference (2012 – 2001)
1	—	120	127	–7
2	—	NA	NA	NA
3	—	NA	122	NA
4	—	126	NA	NA
5	—	NA	123	NA
6	—	120	127	–7
7	—	NA	124	NA
8	—	120	121	–1
9	—	121	124	–3
10	120–125	121	126	–5
11	120–129	121	128	–7
12	120–131	123	132	–9
13	120–135	126	134	–8
14	120–141	131	138	–7
15	127–147	139	142	–3
16	138–151	146	148	–2
17	146–155	151	152	–1
18	149–157	154	155	–1
19	151–159	156	158	–2
20	154–161	158	161	–3
21	156–162	159	162	–3
22	158–165	161	164	–3
23	160–167	164	166	–2
24	163–169	166	168	–2
25	165–172	169	170	–1
26	168–174	172	173	–1
27	171–179	175	175	0
28	174–181	177	177	0
29	175–183	180	180	0
30	177–185	181	182	–1
31	179–187	183	184	–1
32	181–190	185	188	–3
33	182–194	188	191	–3
34	185–199	193	194	–1
35	190–200	198	196	2
36	194–200	199	198	1

Table 6.4: Conditional Average PSAE Mathematics Means, Given Students' WorkKeys *Applied Mathematics* Level Scores

WK Mathematics	PSAE Mathematics 90% Interval	PSAE Mathematics 2012	PSAE Mathematics 2001	Difference (2012 – 2001)
0	120–141	128	126	2
3	125–149	139	139	0
4	138–158	148	148	0
5	147–168	157	158	-1
6	156–179	167	169	-2
7	165–199	180	183	-3

Table 6.5: Conditional Average PSAE Science Means, Given Students' ACT Science Scale Scores

ACT Science	PSAE Science 90% Interval	PSAE Science 2012	PSAE Science 2001	Difference (2012 – 2001)
1	—	129	120	9
2	—	NA	NA	NA
3	—	129	127	2
4	—	128	NA	NA
5	—	127	123	4
6	—	128	125	3
7	126–140	131	127	4
8	125–141	131	127	4
9	126–140	132	129	3
10	127–142	132	130	2
11	127–145	135	132	3
12	128–149	136	134	2
13	129–149	138	136	2
14	130–151	140	139	1
15	131–154	142	142	0
16	135–156	145	144	1
17	137–159	148	148	0
18	139–160	150	152	–2
19	143–164	154	156	–2
20	149–166	158	160	–2
21	152–169	161	163	–2
22	156–172	164	166	–2
23	158–174	166	169	–3
24	162–177	170	173	–3
25	166–178	172	175	–3
26	167–180	175	178	–3
27	170–182	177	180	–3
28	173–184	178	182	–4
29	174–186	180	184	–4
30	176–187	182	183	–1
31	177–190	184	186	–2
32	179–192	186	184	2
33	181–194	188	188	0
34	—	186	186	0
35	184–197	191	190	1
36	186–199	193	193	0

Table 6.6: Conditional Average PSAE Science Means, Given Students' ISBE-Developed Science Scale Scores

ISBE Science	PSAE Science 90% Interval	PSAE Science 2012	PSAE Science 2001	Difference (2012 – 2001)
40	—	NA	122	NA
41	—	123	NA	NA
42	—	129	NA	NA
43	—	126	122	4
44	—	127	NA	NA
45	125–133	127	124	3
46	125–133	128	NA	NA
47	125–133	128	125	3
48	126–134	129	NA	NA
49	126–136	129	126	3
50	126–135	130	127	3
51	127–137	131	NA	NA
52	127–137	131	128	3
53	127–140	132	130	2
54	128–140	133	132	1
55	128–143	134	NA	NA
56	129–143	135	133	2
57	130–146	137	135	2
58	130–148	138	136	2
59	132–150	139	138	1
60	132–151	141	140	1
61	134–153	142	NA	NA
62	135–155	144	143	1
63	136–157	146	144	2
64	138–158	148	146	2
65	140–161	150	148	2
66	141–162	151	151	0
67	142–164	153	153	0
68	143–165	155	155	0
69	146–166	157	157	0
70	147–168	159	NA	NA
71	150–170	160	159	1
72	151–173	162	162	0
73	152–173	163	164	–1
74	154–175	165	NA	NA
75	156–177	166	166	0
76	157–177	168	168	0
77	158–178	169	NA	NA
78	160–179	171	171	0
79	161–181	172	173	–1
80	162–181	173	NA	NA
81	164–184	175	175	0
82	165–184	175	NA	NA
83	166–185	177	177	0
84	167–186	178	NA	NA
85	168–187	179	180	–1

Table 6.6: Conditional Average PSAE Science Means, Given Students' ISBE-Developed Science Scale Scores

ISBE Science	PSAE Science 90% Interval	PSAE Science 2012	PSAE Science 2001	Difference (2012 – 2001)
86	170–188	180	NA	NA
87	170–188	181	NA	NA
88	173–190	182	182	0
89	175–192	184	NA	NA
90	175–192	184	NA	NA
91	176–194	186	185	1
92	177–195	187	NA	NA
93	177–195	187	NA	NA
94	178–195	189	NA	NA
95	180–197	190	187	3
96	182–199	192	NA	NA
97	181–199	192	NA	NA
98	—	NA	NA	NA
99	184–200	195	NA	NA
100	—	198	191	7

Chapter 7
Quality Control Procedures for Scoring, Analysis, and Reporting

Introduction

Quality control procedures have been established to ensure that all PSAE materials are accurately, efficiently, and reliably developed, produced and scored. Facilities, personnel, equipment, processes, procedures, safeguards have been put in place to ensure that all materials including answer documents, test materials, and administration materials are handled securely.

Established quality assurance verification and validation procedures are executed throughout all PSAE development, and are meticulously continued throughout the duration of the PSAE processing procedures. Established industry standard quality control procedures are described in this chapter regarding processes such as scoring, quality control checks, verifying analyses, checking output from scoring programs (to ensure accuracy), and reporting.

Quality assurance and control begins at the earliest possible stage (including planning meetings with ISBE and ACT) and continues throughout reviews, advanced quality planning, process controls, inspections and testing, to final delivery of reports. Each production area has several quality control checks and control methods—including inspections and system verifications and validations—built into the standard procedures. Refined validity checks, scanner accuracy checks, editing procedures, error corrections, and other quality controls result in maximum accuracy in reported results. These combined assurances result in an accurate collection of data for scoring, analysis, and reporting.

Initial Steps

Student enrollment and demographic data are gathered prior to test administration allowing for efficient production of test booklets, shipping materials, and initial file layouts for reports. Test booklets are serialized to ensure accountability from their creation, throughout shipping, receipt, test administration, post-test packaging and shipping, through final storage. All report requirements are established prior to test administration. Samples of reports are generated and must be approved by ISBE prior to their publication.

Prior to Scoring, Reporting Processes Verified

In order to maintain accurate reporting of results, reports are generated from test data and from live data. Comparing these reports provides the opportunity to identify discrepancies between expected results and actual report results. Several test cases are executed in order to check accuracy prior to distribution of results. Test cases are constructed to check varying combinations of districts, schools, and grades. Individual and summary reports are tested. Report formats are compared with input sources of approved samples. Student data are validated and verified by querying the appropriate student data. Batches from first production are collated and analyzed to validate all processes are running correctly.

Scoring

Both technological and human quality control measures are used to ensure accurate scoring. Technologically speaking, the scanning equipment is highly sensitive to the presence or absence of a mark in the areas of the answer document thus allowing for detection of potential erasures, double-grids, and excessive or suspicious patterns in responses. Summary reports of these identified actions are analyzed and made available for validation and follow-up actions.

Several additional quality control procedures are executed by staff members in order to monitor and control the accuracy of the scoring process. One out of every 100 documents is hand-scored by staff throughout the entire scoring process to ensure accuracy. Experienced psychometric staff members perform empirical reviews of the preliminary scoring results for each and every item from early samples from the administration. Although answer keys undergo several reviews for accuracy throughout the development process, this last empirical review is designed to identify the possibility of an incorrect scoring key and to raise questions about poorly performing items. These preliminary analyses are performed on early materials in sufficient time to adjust the keys if required prior to scoring. Consensus regarding all correct answers is required before official scoring is allowed to begin.

Analyses

Once scoring is underway, several analyses are executed to ensure the accuracy and reasonableness of results. Established file-naming conventions are in place to assure that processes such as equating, scaling, calibration checks, DIF and item analyses are executed accurately using appropriate data files. Established step-by-step procedures across departments are followed within given timelines to assure each area gets sufficient time to rigorously run all tests, reports, and rechecks of analyses.

Reporting

Multiple quality control procedures are in place to ensure that all PSAE results are correctly attributed to the students, school, districts, and/or other subgroups for whom aggregate assessment results are requested. Bar-coding of all secure test materials provides for accurate accountability from their creation through final storage and eventual disposal. Test booklets are serialized to provide additional accountability for each student, assuring that scanned scores are correctly attributed to appropriate students. Test reports developed are checked to assure accuracy of information reported. Even mailing labels undergo quality assurance checks to make sure that reports are mailed to the proper location.

Chapter 8
Results of the 2012
Prairie State Achievement Examination

This chapter provides a summary of the results of the Spring 2012 PSAE administration. Individual and school PSAE reports from the 2012 administration were shipped to schools in September 2012. The PSAE Goals Reports for individual students and for schools were also shipped in September 2012. In addition to the PSAE reports, individual WorkKeys score reports for *Reading for Information* and *Applied Mathematics* were shipped to schools in August 2012 for distribution to students. Individual ACT reports had been mailed in May and June 2012 to students at their homes, along with ACT's standard student guide for interpreting scores. Home high schools also receive a copy of each student's ACT score report. Students receive a Prairie State Achievement Award for any PSAE score or scores in the Exceeds Standards performance level.

PSAE Score Results

Approximately 147,559 students sat for the spring administration of the PSAE test battery in April and May 2012, although not all students took the full battery of tests. Table 8.1 shows the average score for the state for each of the three PSAE subject tests, and the state average for the component assessments that make up each PSAE subject test.

Table 8.2 shows the percentage of students in each of the four performance levels for the state for each of the three PSAE subject tests. The percentage of students meeting or exceeding standards ranged from 51% to 52%, compared to 49% to 54% reported for spring 2011.

Table 8.3 contains the percentage of students in each of the four performance levels by PSAE subject; scores are disaggregated by gender, ethnicity, income level, disability, and migrant status. Results are provided only if five or more students are present in a given category. Ethnicity categories were changed in 2011 to parallel federal guidelines for reporting ethnicity. The results in 2012 are similar to those reported in 2011.

Table 8.1: Average PSAE Scores for Grade 11 Students

PSAE test	Score range	Average score
PSAE Reading	120–200	155
ACT Reading	1–36	19
WorkKeys *Reading for Information*	<3, 3–7	5
PSAE Mathematics	120–200	156
ACT Mathematics	1–36	20
WorkKeys *Applied Mathematics*	<3, 3–7	5
PSAE Science	120–200	158
ACT Science	1–36	20
ISBE-Developed Science	40–100	70
ACT English	1–36	20

Table 8.2: Percentage of Grade 11 Students in Each of the Four PSAE Performance Levels

	Performance levels				
PSAE scores	Academic Warning	Below Standards	Meets Standards	Exceeds Standards	Meets or Exceeds Standards*
Reading	10%	39%	42%	9%	51%
Mathematics	11%	38%	42%	9%	52%
Science	9%	40%	41%	11%	52%

Note: Due to rounding, percentages may not sum to 100.
*May not equal the sum of the two previous columns due to rounding.

Table 8.3: Percentage of Grade 11 Student Scores Within Each PSAE Performance Level by Various Categories

	Reading				Mathematics				Science			
	Academic Warning	Below	Meets	Exceeds	Academic Warning	Below	Meets	Exceeds	Academic Warning	Below	Meets	Exceeds
All students	10	39	42	9	11	38	42	9	9	40	41	11
Female	8	39	44	10	10	40	42	8	9	43	41	8
Male	12	40	40	8	11	35	42	11	9	36	41	13
Hispanic or Latino	15	52	31	3	14	50	34	3	12	55	30	3
American Indian or Alaska Native	11	43	40	6	10	41	44	5	9	48	37	6
Asian	7	27	49	17	4	19	49	29	5	23	49	24
Black or African American	19	57	23	1	25	54	20	1	21	60	18	1
Native Hawaiian or Other Pacific Islander	8	42	37	12	10	35	47	8	9	39	43	8
White	6	31	51	13	6	30	52	13	4	29	51	15
Two or More Races	7	37	44	11	9	37	43	11	7	37	43	13
Low income	17	52	28	2	19	51	28	2	16	55	26	2
Not low income	5	31	51	14	5	29	52	15	4	29	51	16
LEP	57	41	3	0	42	46	10	1	43	51	6	0
Non-LEP	9	39	42	9	10	38	43	10	8	39	42	11
IEP	41	43	14	2	46	41	12	1	37	48	13	2
Non-IEP	6	39	45	10	6	37	46	10	5	38	45	12
Migrant	54	42	4	0	46	50	4	0	31	65	4	0
Non-migrant	10	39	42	9	11	38	42	9	9	40	41	11

Note: Due to rounding, percentages may not sum to 100.

PSAE Trend Data

Tables 8.4, 8.5, and 8.6 contain scale score summary statistics for the the PSAE subject areas for the spring administrations in 2012 (three subject areas), and in 2011 and 2010 (four subject areas). All forms and all students with scores are included. As can be seen from the tables, the sample sizes increase by about 4,000 from 2010 to 2011 and then level off from 2011 to 2012. The means for Reading are fairly steady across the three years as are the Reading standard deviations. The means for Mathematics are nearly the same for the three years, but the standard deviations for Mathematics display a little variably across the three years. The Science means are stable from 2010 to 2011 but increase in 2012; the Science standard deviations have a small dip in value for 2011. The Writing means and standard deviations are comparable for 2010 and 2011. The ACT essay was not administered in 2012 so there are no Writing test statistics for that year.

Although the means and standard deviations for all three subjects are very stable across the three years, there is some slight variation from year to year, which is likely statistically significant because of the large sample sizes. However, the practical significance of this variation when compared to the size of the subject standard deviations is not great. Even a mean difference of 1 point from year to year is not very large when divided by a standard deviation of 16 which is the usual method for determining the practical effect size of mean differences.

The percent Meets/Exceeds column represents the percentage of examinees that received either a meets or exceeds level score in the specified subject. The percent Meets/Exceeds for Reading tends to decrease over the three years, but the percent Meets/Exceeds for Mathematics increases over the three years. The Science percent Meets/Exceeds dips about one percentage point from 2010 to 2011 and then increases over 2.5 percentage points from 2011 to 2012. The Writing percent Meets/Exceeds increases slightly from 2010 to 2011. The PSAE scale score distributions are unimodal and only slightly skewed which means most of the scores fall in the middle of the distribution near the meets category cut-score, so small shifts in the shape of the distribution near the meets cut-score from year to year can have large effects on the percent Meets/Exceeds. That is because scores near the center of the distribution have large numbers of students, a small shift in the scale of a point or two near a cut-score can affect many students. This could help explain the changes in the percent Meets/Exceeds statistics over the years.

Table 8.7 presents the correlations among the three 2012 PSAE scores. The correlations are fairly homogenous with an average value of about 0.83 and a range of about 0.78 to 0.87. This homogeneity among the correlations suggests that one component can explain most of the variance among the three tests. Tables 8.8 and 8.9 present the results of a principal component analysis of the correlation matrix for the three tests. Table 8.8 contains the eigenvalues and the proportion of variance explained for each principal component. The first principal component has an eigenvalue of 2.65 and accounts for about 88% of the variance among the three tests. The remaining components all have eigenvalues less than one, and combined only account for about 12% of the variability. This further indicates a one component model fits the data well. Table 8.9 contains the loadings of the three tests on the first principal component. All three tests load nearly equally and very highly on the first principal component. This indicates that students tend to perform the same, either well or poorly, on all three tests rather than perform differently on different tests.

Figures 8.1, 8.2, and 8.3 show the percentages of students who meet or exceed the Illinois Learning Standards on 0, 1, 2, or 3 PSAE Tests for different groups. Figure 8.1 gives the percentages for the entire group of students, Figure 8.2 gives the percentages for males and females separately, and Figure 8.3 gives the percentages for different ethnic groups.

Table 8.4: PSAE Spring 2012 Scale Score Summary Statistics—All Forms Included

Subject	N	Mean	SD	Variance	Skewness	Kurtosis	% Meets/Exceeds
Reading	145,256	154.9300	15.7384	247.6976	0.1125	-0.5521	50.69
Mathematics	145,377	156.3833	16.3441	267.1281	0.0925	-0.0548	51.62
Science	145,348	157.8408	15.5373	241.4074	0.0000	-0.8147	51.67

Table 8.5: PSAE Spring 2011 Scale Score Summary Statistics—All Forms Included

Subject	N	Mean	SD	Variance	Skewness	Kurtosis	% Meets/Exceeds
Reading	145,468	155.5119	16.0110	256.3509	0.0549	–0.5333	51.02
Mathematics	145,565	156.0707	16.1977	262.3662	0.1004	–0.0233	51.30
Science	145,559	157.0752	15.0184	225.5528	0.0376	–0.7816	49.19
Writing	146,044	156.2842	16.4922	271.9937	–0.1617	–0.3778	53.71

Table 8.6: PSAE Spring 2010 Scale Score Summary Statistics—All Forms Included

Subject	N	Mean	SD	Variance	Skewness	Kurtosis	% Meets/Exceeds
Reading	141,316	155.6154	16.1664	261.3540	0.0532	–0.5796	52.00
Mathematics	141,416	156.3900	17.1942	295.6390	0.1600	–0.1542	50.40
Science	141,382	156.9900	15.6490	244.8924	0.0271	–0.8270	50.17
Writing	142,025	155.7972	16.3981	268.8987	–0.1406	–0.4136	52.95

Table 8.7: Correlations Among 2012 PSAE Scores

	Reading	Mathematics	Science
Reading	1.00000	0.77901	0.83293
Mathematics	0.77901	1.00000	0.86672
Science	0.83293	0.86672	1.00000

N = 145,220

Table 8.8: Eigenvalues of the Correlation Matrix

Component	Eigenvalue	Difference	Proportion	Cumulative
1	2.65296883	2.42827491	0.8843	0.8843
2	0.22469392	0.10235666	0.0749	0.9592
3	0.12233726		0.0408	1.0000

Table 8.9: First Principal Component Loading Values Across Years

PSAE area	First principle component loadings											
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Reading	.91	.92	.92	.92	.93	.93	.92	.92	.93	.92	.93	.92
Mathematics	.91	.91	.91	.91	.94	.94	.92	.92	.93	.93	.93	.94
Science	.94	.95	.95	.94	.96	.95	.94	.94	.94	.94	.94	.96
Writing							.89	.91	.90	.91	.92	

Figure 8.1: Percentage of Students Achieving "Meets Standards" or Above for PSAE Spring 2012

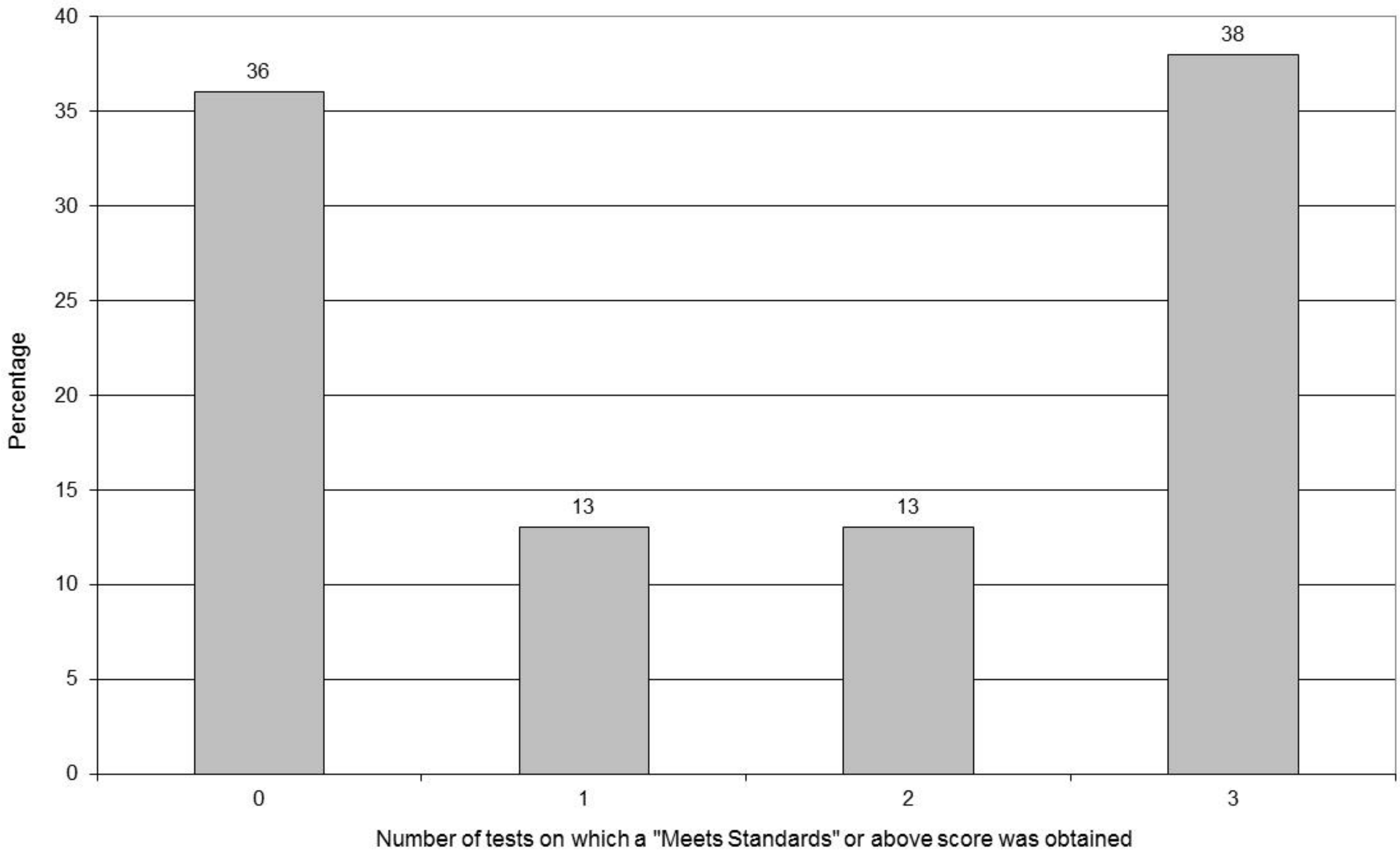


Figure 8.2: Percentage of Students Achieving "Meets Standards" or Above by Gender for PSAE Spring 2012

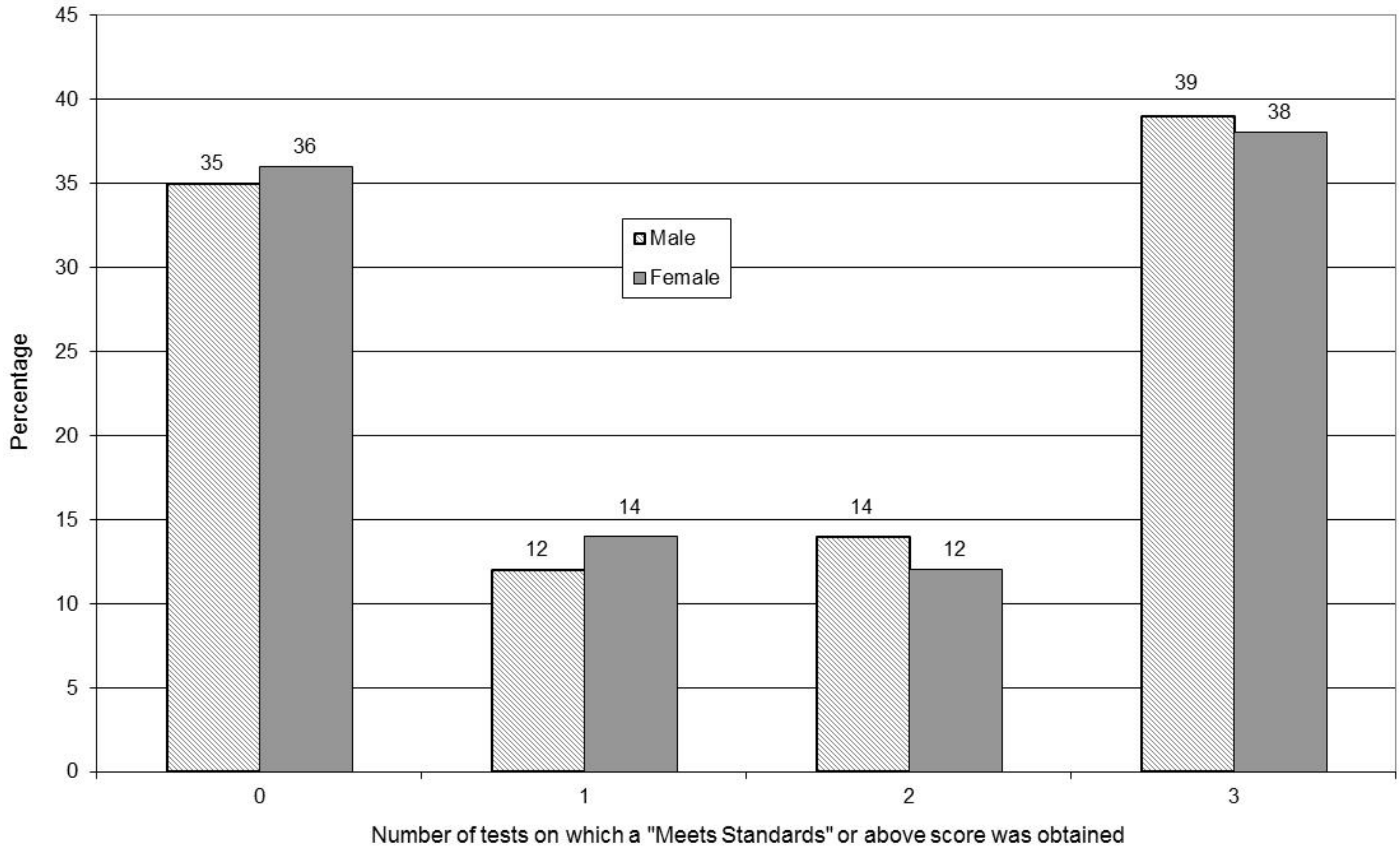
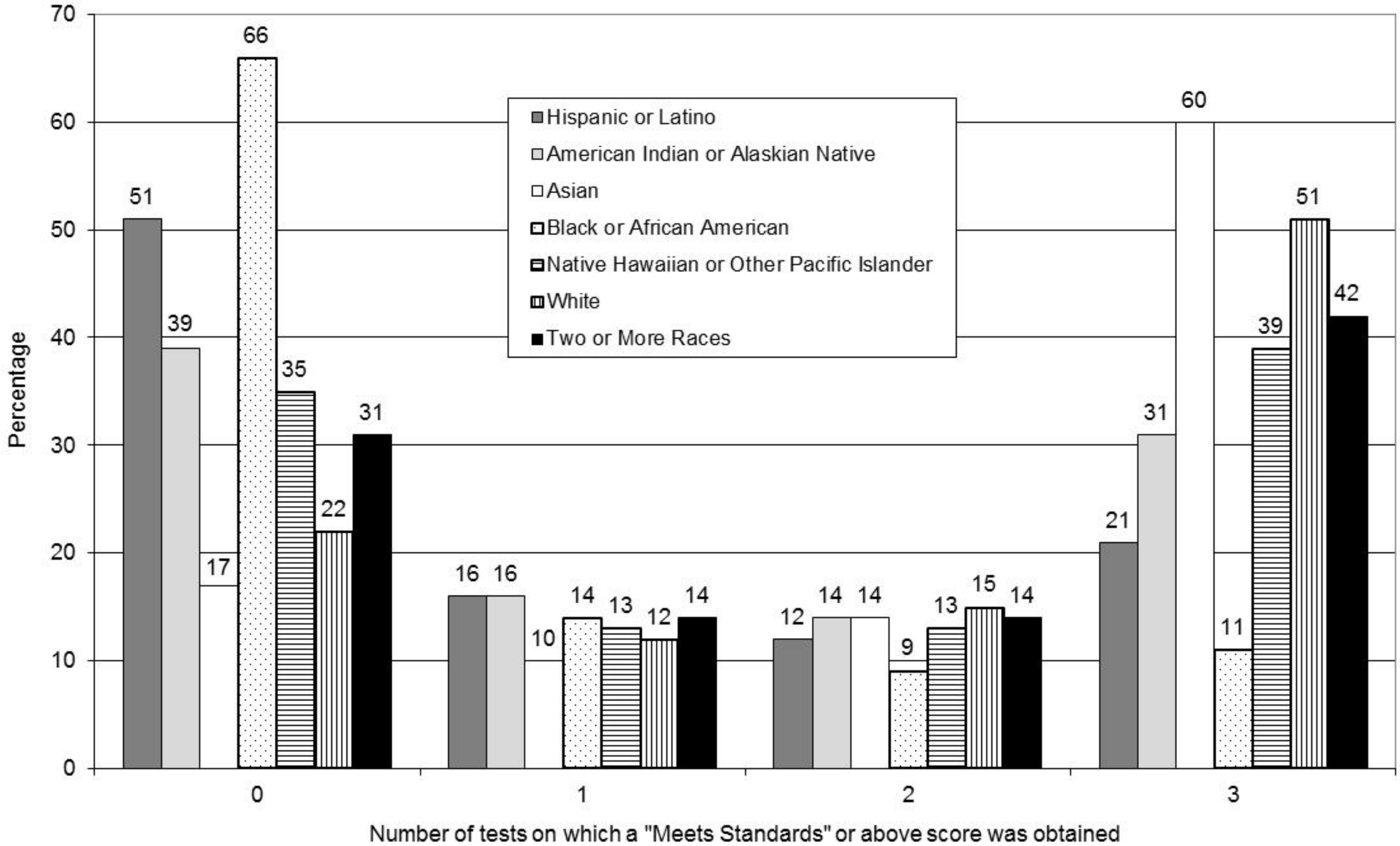


Figure 8.3: Percentage of Students Achieving “Meets Standards” or Above by Ethnicity for PSAE Spring 2012



Chapter 9
Illinois State Goals Reports

The Illinois State Goals reports provide information about students' PSAE performance by State Goals in English Language Arts, Mathematics, and Science.

The student report provides information regarding a student's strengths and weaknesses relative to the Illinois State Goals assessed by the PSAE. The report shows 1) the total number of essay points and/or test questions on the PSAE based on each State Goal, 2) the number of essay points received and/or test questions a student answered correctly for each State Goal, and 3) the number of essay points and/or test questions a typical student who performed at the "Meets Standards" level in a given content area received and/or answered correctly.

The school report provides the number or range of number of test questions for each State Goal and the average percent correct for the school, the district, and the state based on multiple-choice test questions only. The school report also includes a description of each State Goal and the component tests that contribute to each of the three PSAE subject scores.

The 2012 administration state percent correct results in each PSAE subject area are shown in Table 9.1 below.

Table 9.1: 2012 State Percent Correct by PSAE Subject Area

PSAE Component		State Goal	Standard(s)	Number/Range of Questions	State Percent Correct
Reading	1:	Vocabulary Development, Reading Strategies, and Reading Comprehension	1A, 1B, 1C	70	60.9%
Mathematics	6:	Number Sense	6A, 6B, 6C, 6D	25–33	69.7%
	7:	Measurement	7A, 7B, 7C	13–19	45.9%
	8:	Algebra	8A, 8B, 8C, 8D	20–28	51.6%
	9:	Geometry	9A, 9B, 9C, 9D	11–19	44.4%
	10:	Data Analysis, Statistics, and Probability	10A, 10B, 10C	4–12	63.3%
Science	11:	Scientific Inquiry and Technological Design	11A, 11B	42	53.3%
	12:	Life Sciences and Environmental Sciences Matter, Energy, and Forces Earth and Space Sciences	12A, 12B 12C, 12D 12E, 12F	30	56.5%
	13:	Safety, Practices of Science, Science/ Technology/Society, and Measurement	13A, 13B	8	65.7%

References

ACT. (1999). *Comparison of the Illinois Learning Standards to the ACT Assessment, PLAN, and EXPLORE*. Iowa City, IA: Author.

ACT. (2000). *Comparison of the Illinois Learning Standards to the ACT Assessment Standards for Transition*. Iowa City, IA: Author.

ACT. (2006). *Comparison of the Illinois Learning Standards to the ACT Assessment, PLAN, and EXPLORE*. Iowa City, IA: Author.

ACT. (2007). *ACT technical manual*. Iowa City, IA: Author.

ACT. (2008a). *WorkKeys Applied Mathematics technical manual*. Iowa City, IA: Author.

ACT. (2008b). *WorkKeys Reading for Information technical manual*. Iowa City, IA: Author.

ACT. (2009). *ACT National Curriculum Survey® 2009*. Iowa City, IA: Author.

AERA. *See* American Educational Research Association, American Psychological Association, National Council on Measurement in Education.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory* (pp. 68–83). New York: Holt, Rinehart, and Winston.

Gulliksen, H. (1987). *Theory of mental tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. A. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction* (pp. 60–90). Princeton: Princeton University Press.

Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement, 27*, 345–359.

Illinois State Board of Education. (2001). *Prairie State Achievement Examination Technical Manual*. Iowa City, IA: ACT, Inc.

Illinois State Board of Education. (2002). *Prairie State Achievement Examination Technical Manual*. Iowa City, IA: ACT, Inc.

Illinois State Board of Education. (2003). *Prairie State Achievement Examination Technical Manual*. Iowa City, IA: ACT, Inc.

Illinois State Board of Education. (2004). *Prairie State Achievement Examination Technical Manual*. Iowa City, IA: ACT, Inc.

Illinois State Board of Education. (2005). *Prairie State Achievement Examination Technical Manual*. Iowa City, IA: ACT, Inc.

Illinois State Board of Education. (2007). *Prairie State Achievement Examination Technical Manual*. Iowa City, IA: ACT, Inc.

Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics* (4th ed., Vol. 1). New York: Macmillan.

Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics*, *9*, 25–44.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.

Kolen, M. J., Zeng, L., & Hanson, B. A. (1996). Conditional standard errors of measurement for scales scores using IRT. *Journal of Educational Measurement*, *33*, 129–140.

Lee, W. (2010). Classification consistency and accuracy for complex assessments using item response theory. *Journal of Educational Measurement, 47*, 1–17.

Lee, W., Brennan, R. L., & Hanson, B. A. (2000). *Procedures for computing classification consistency and accuracy indices with multiple categories*. ACT Research Report Series 2000-10. Iowa City, IA: ACT.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*, 179–197.

Lord, F. M. (1980). *Application of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Millman, J., & Greene, J. (1989). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: American Council on Education and Macmillan Publishing Company.

Mislevy, R. J., & Bock, R. D. (1990). *BILOG 3. Item analysis and test scoring with binary logistic models* (2nd ed.). Mooresville, IN: Scientific Software.

Schulz, E. M., Kolen, M. J., & Nicewander, W. A. (1997). *A study of modified-Guttman and IRT-based level scoring procedures for WorkKeys assessments.* ACT Research Report Series 97-7. Iowa City, IA: ACT.

Schulz, E. M., Kolen, M. J., & Nicewander, W. A. (1999). A rationale for defining achievement levels using IRT-estimated domain scores. *Applied Psychological Measurement, 23*, 347–362.

Spray, J. A. (1989). *Performance of three conditional DIF statistics in detecting differential item functioning on simulated tests*. (ACT Research Report No. 89-7). Iowa City, IA: ACT.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Journal of Applied Psychological Measurement, 7(2)*, 201–210.

Thorndike, R. L. (1951). Reliability. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 560–620). Washington, DC: American Council on Education.

Webb, N. L. (2006a, March). *Alignment analysis of mathematics goals and assessments: Illinois grade 11.* (Available from the Illinois State Board of Education, 100 N. First St., Springfield, IL 62777)

Webb, N. L. (2006b, March). *Alignment analysis of reading goals and assessments: Illinois grade 11.* (Available from the Illinois State Board of Education, 100 N. First St., Springfield, IL 62777)

Webb, N. L. (2006c, April). *Alignment analysis of science goals and assessments: Illinois grade 11.* (Available from the Illinois State Board of Education, 100 N. First St., Springfield, IL 62777)

White, E. M. (1994). *Teaching and assessing writing: Recent advances in understanding, evaluating, and improving student performance.* San Francisco: Jossey-Bass.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.

Yen, W. M. (1983). Tau-equivalence and equipercentile equating. *Psychometrika*, *48*, 353–369.

Appendix A

Procedures for Requesting ACT Test Accommodations for Day 1 of the Prairie State Achievement Examination, Spring 2012

ACT

Procedures for Applying for ACT Test Accommodations – Spring 2012
Prairie State Achievement Examination (PSAE)

Your ACT-Approved Accommodations receipt deadline is **Friday, January 27, 2012**.

Your State-Allowed Accommodations receipt deadline is **Friday, March 9, 2012**.

OVERVIEW

ACT provides test accommodations in accordance with *Title III of the Americans with Disabilities Act (ADA*). Schools provide accommodations under different regulations. Thus, having a diagnosis and receiving accommodations in school do not guarantee ACT approval of those accommodations.

Note: If the Illinois State Board of Education (ISBE) chose, as part of their contract with ACT, the option of Plus Writing, that information is easily identified throughout these procedures as Writing Test only. ISBE recommends that schools apply for writing accommodations for students as applicable in case the ACT Writing test is reinstated for this year.

You may photocopy the accompanying accommodations forms or download forms from the PSAE website: http://www.isbe.net/assessment/psae.htm. Two different forms are available to request accommodations for the ACT (PSAE Day 1). Review the information below to determine which one of these forms to complete for each student.

1. ACT-Approved Accommodations. Complete an *Application for ACT-Approved Test Accommodations* for students with diagnosed disabilities who are receiving special education services described in a current Individualized Education Program (IEP) or Section 504 Plan. Only students who have an IEP or Section 504 Plan are eligible to apply for ACT-Approved Accommodations. ACT-Approved Accommodations are not available for students solely on the basis of limited English proficiency. **ACT must receive the forms by the ACT-Approved Accommodations receipt deadline of Friday, January 27, 2012.**
2. State-Allowed Accommodations. Students who do not meet the eligibility requirements stated below (or whose application for ACT-Approved Accommodations is denied or additional accommodations are needed beyond what was approved by ACT) may request State-Allowed Accommodations by submitting a *Request for State-Allowed Accommodations* by the **State-Allowed Accommodations receipt deadline of Friday, March 9, 2012.** Scores earned with State-Allowed Accommodations will be used for PSAE purposes but will **not** be reported by ACT to colleges, scholarships agencies, or any other entities.

ELIGIBILITY REQUIREMENTS

To be considered for ACT-Approved Accommodations, students must meet **ALL** of the following requirements:

1. Professionally Diagnosed Disability. The student's disability must have been professionally diagnosed by a qualified professional, or team of professionals, whose credentials are appropriate to the disability. Current written diagnostic documentation of the disability must be on file at school and must meet **ALL** the "Guidelines for Documentation" listed in this document.
   - If diagnosed for the FIRST time before September 2008, reconfirmation is required within the last 3 years. A current IEP or 504 Plan on file at the school may serve as reconfirmation, provided the initial diagnosis was made by a qualified professional(s).
   - If FIRST diagnosed within the last 3 years, submit complete written diagnostic documentation *with* the application form.
2. The current IEP or 504 Plan must document ALL accommodations applied for are provided in school. Submit a copy of the student's *current* IEP or *current* 504 Plan that supports the need for all requested accommodations due to the disability. The student's name and effective dates of the IEP or 504 Plan must appear on all pages submitted. If the plan has been in place less than 3 years, full diagnostic documentation is required.

AUTHORIZATION TO PROVIDE CONFIDENTIAL DOCUMENTATION

Schools are required to provide the necessary information and documentation to support applications for test accommodations. ISBE has authorized ACT to collect and review this documentation. All documentation provided to ACT will be kept confidential and will not become part of the student's ACT score record.

EXAMPLES OF TEST ACCOMMODATIONS

Extended Time and/or Alternate Formats. If the student's professionally diagnosed and documented disability requires one or more of the accommodations below, the school must submit a completed application form:

- more than standard time
- testing over multiple days
- additional or stop-the-clock breaks
- alternate test formats such as Braille, cassettes or DVDs, or a reader, and/or alternate response modes
- Large type test booklet: If a student requires a large type test booklet (18-point) but can test with standard time limits (including the standard break(s) allowed), the school must submit a completed application form specifying the accommodations requested. Refer to Section E on the application.
- Writing Test only: Use of a scribe or computer for the Writing Test (typically for disabilities that prevent students from writing independently).
- Writing Test only: Extended time on the Writing Test only (students with developmental writing disorder, written expression, or dysgraphia).

Local Decision Accommodations. If a student can test in a single session with standard time limits (including the standard break(s) allowed) and use a regular (10-point) test booklet, but the disability requires other accommodations, **the school may make such arrangements without prior consultation with ACT**. Examples include, but are not limited to:

- assignment to a wheelchair accessible room,
- permission for diabetics to eat snacks (requires a separate room),
- permission to use Irlen filters or color overlays,
- marking answers in test booklet (no extended time).

Examples of accommodations for students with hearing impairments that do not require time extensions or ACT approval include:

- a sign language interpreter (not a relative) to sign all spoken instructions (not the test items),
- seating near the front of the room to lipread spoken instructions,
- a written copy of spoken instructions with visual notification from testing staff of test start, five minutes remaining, and stop times.

Deadline for Submitting Applications

To be considered for testing, applications and all required documentation for ACT-Approved Accommodations must be **received by ACT no later than the ACT-Approved Accommodations deadline of Friday, January 27, 2012**. The Test Accommodations Coordinator at each school is responsible for gathering all completed forms and submitting them as a group to arrive by the deadline *with* a **completed** Test Accommodations Coordinator Header to:

ACT State Test Accommodations
301 ACT Drive
PO Box 4071
Iowa City, IA 52243

Instructions for Completing the Application Form and Providing Required Documentation

A school official such as a counselor, special education teacher, or principal is to complete a form for **each** student for whom accommodations are requested. **The form may be photocopied as needed.** The application form will **not** be processed if it is missing required information or signatures or is received after the deadline. If any of the information provided is false, ACT reserves the right to cancel scores.

Note: The most frequent reason accommodations are not approved is incomplete or inaccurate information on the application form. Please read and follow instructions carefully.

Side 1 of the *Application for ACT-Approved Test Accommodations* form

A. **Student Information.** Student address is required. If not available, school address may be used.

B. **Previous Approval of the Same Accommodations on the ACT.** Mark the appropriate answer. If no, complete both sides of the application and submit required documentation.

C. **Diagnosed Disability.** Check all applicable disabilities as stated in written documentation on file at the school. Pay attention to those diagnoses that require full documentation for approval. Include FSIQ where requested. <u>Writing Test only:</u> Students with developmental writing disorder, written expression, or dysgraphia apply for extended time on the Writing Test only.

D. **Test Format Requested.** The type of materials applied for must be supported by the accommodations plan at school or on a previous "ACT Accommodations Approval" letter for this student. Documentation of a visual disability is required to support requests for large type test booklets. Both scannable and large block answer sheets are provided with each large type booklet. If no test format is selected, regular type will be assigned.
Important—Students using cassettes/DVDs may test as a group. Students must use their own headphones and begin each test at the same time. Before requesting DVDs, work with technical personnel at your school. Order the practice ACT tests on DVDs so that you can test them on your equipment. Also have students take the practice tests so they will be comfortable using DVDs on test day. ACT will provide usage guidelines and track listings with each set of DVDs.

E. **Time Requested.** Mark the option most similar to the accommodations normally provided at school. ACT will assign a timing code based on the disability and approved test format.

F. **Other Accommodations Requested.** If needed due to the disability, explain in detail and submit supporting documentation. Complete only if other accommodations are requested. <u>Writing Test only:</u> If a student is unable to produce independent writing, list the accommodations normally provided at school for writing tests (e.g., use of a scribe or computer).

Side 2 of the *Application for ACT-Approved Test Accommodations* form

G. **Specific Disorder or Condition**. Must be specific. Follow the instructions on the form. The following terms are **not** sufficiently specific: specific learning disabilities (SLD), other health impaired, perceptual communication disorder, processing disorder, etc. For learning disabilities, please use the DSM-IV diagnosis, if available, as stated on the documentation from the diagnosing professional.

H. **History of Diagnosis**. The diagnosing professional's credentials must be appropriate to the disability. If the disability was identified by an IEP team, list relevant titles and specializations.

H-a. If **FIRST** diagnosed <u>*before* grade 9</u>, complete only the "age or grade of student" when diagnosed. **If FIRST diagnosis was within the last 3 years, submit complete diagnostic documentation *with* the application form** (see "Guidelines for Documentation" section).

H-b. **If recently re-confirmed, there must be a re-confirmation within the last 3 years** by a psychologist, learning disabilities specialist/team, or other qualified professional, or team of professionals, with direct knowledge of the student's disability. A current IEP or 504 Plan on file at the school may serve as reconfirmation.

I. **Current IEP or 504 Plan on File at School.** Indicate the type of accommodations plan now on file at the school, and attach the required copy. The student's name and effective dates of the IEP or 504 Plan must appear on all submitted pages.

J. **School Official's Signature.** Read and sign the statement. A relative of the student may *not* sign.

K. **Student/Parent Signature.** *If the student is 18 or older, the student must sign. If the student is younger than 18, his/her parent or legal guardian must sign.* **School official may sign for the parent if approval has been obtained by phone; note "per phone call" and initial.**

ACT Guidelines for Documentation

Documentation must be written by the diagnosing professional and must meet **ALL** of these guidelines:

1. **States the specific impairment** as diagnosed
2. **Is current** (no older than September 2008)
3. **Describes presenting problem(s) and developmental history**, including relevant educational and medical history
4. **Describes the comprehensive assessments** (neuropsychological or psychoeducational evaluations), including evaluation dates, used to arrive at the diagnosis:
   - *For learning disabilities*, must provide test results (including subtests), **with standard scores** and percentiles, from
     a) an aptitude assessment using a complete, valid, and comprehensive battery,
     b) a complete achievement battery,
     c) an assessment of information processing, **and**
     d) evidence that alternative explanations were ruled out.
   - *For ADD/ADHD*, must include
     a) evidence of early impairment,
     b) evidence of current impairment, including presenting problem and diagnostic interview,
     c) evidence that alternative explanations were ruled out,
     d) results from valid, standardized, age-appropriate assessments, **and**
     e) number of applicable DSM-IV criteria and description of how they impair the individual.
   - *For visual, hearing, psychological, emotional, or physical disorders*, provide detailed results from complete ocular, audiologic, or other appropriate diagnostic examination.
5. **Describes the substantial limitations (e.g., adverse effects on learning, academic achievement, or other major life activities) resulting from the impairment**, as supported by the test results
6. **Describes specific recommended accommodations** and provides a rationale explaining how these specific accommodations address the substantial limitations
7. **Establishes the professional credentials of the evaluator**, including information about licensure or certification, education, and area of specialization

Complete details about ACT's policies for documentation of requests for test accommodations are available at: **www.act.org/aap/disab/policy.html**

RELATED INFORMATION

Application forms are processed in the order they are received at ACT.

1. Notification of PSAE Day 1 Accommodations Approval. ACT will send an authorized accommodations letter for each student approved for accommodations on the ACT to the school's Test Accommodations Coordinator. If a student's *Application for ACT-Approved Accommodations* is not approved, ACT will send written notification to the Test Accommodations Coordinator, with these options:
   a. Submit Additional Documentation. ACT will notify the Test Accommodations Coordinator at the school if additional documentation is needed to support the student's application. Documentation must be submitted in writing – a fax reply will assist in meeting deadlines. Refer to your *Checklist of Dates* for the deadline to submit this additional documentation.
   b. Request State-Allowed Accommodations. If a student's *Application for ACT-Approved Accommodations* is not approved, you may request State-Allowed Accommodations instead, using the *Request for State-Allowed Accommodations* form. Refer to your *Checklist of Dates* for the deadline to submit this request. If a request to test with State-Allowed Accommodations is not received for a student, the student **is expected to** test with standard time limits and use a regular type (10-point) test booklet *without* accommodations.
   c. Test Standard Time. If you fail to submit additional documentation when it is requested and by the deadline you are given, the student **may** test with standard time limits and use a regular type (10-point) test booklet *without* accommodations or may request State-Allowed accommodations.
2. Determining Day 2 Accommodations. ACT's approval of accommodations applies to PSAE Day 1 administration of the ACT only. However, schools will need to order from Pearson's *PSAE TestSites Online* system the quantity and type of alternate formats needed for the PSAE Day 2 administration.
3. Assignment of Test Materials. ACT assigns specific test materials (by serial number) to each student for PSAE Day 1. **Only the authorized student may use the materials; they may not be used by another student, or transferred to another test site. If ACT procedures are not followed, the resulting scores will be cancelled.** PSAE Day 2 accommodations materials ordered are not assigned to specific students.
4. Timing Codes. ACT will provide specific instructions and will assign a timing code to each student approved for ACT-Approved Accommodations. **Students with different timing codes may not test in the same room; students approved for a reader's script must test individually; and ACT-Approved Accommodations must be administered separate from State-Allowed Accommodations.** You will receive an ACT-Approved Accommodations roster and a State-Allowed Accommodations roster. Do NOT mix these two groups in a room together. If ACT procedures are not followed, the resulting scores may be cancelled.
5. Sequence of the Tests. A student must complete all PSAE Day 1 tests before beginning PSAE Day 2 tests. If ACT procedures are not followed, the resulting scores may be cancelled.
6. Preparing for the ACT. A copy of *Preparing for the ACT,* which includes information about the tests, test-taking strategies, and complete practice tests, is available. Schools have a supply of this free regular type booklet for distribution to students. Many schools have previously ordered a copy of a practice test in Braille, large type, or on cassettes or DVDs for their libraries. If your school does not have copies available, you may order library copies of these alternate format practice tests directly from ACT at no charge. Refer to ACT's website on Services for Students Disabilities at **www.act.org/aap/disab/index.html** for more information. You will receive *Preparing for the ACT Special Testing* with each alternate format ordered; it contains the scoring keys.

ACT Repeat Testing

Students approved for accommodations may, at their option, apply to take the ACT again with the same approved accommodations.

1. <u>During Spring 2012.</u> ACT has adjusted its usual 60-day ACT retest restriction for the state administration. Grade 11 students who wish to take the ACT with extended time more than once during the spring may do so, as follows:
   - Students who can test with regular type or large type materials with up to 50 percent additional time may request to test with accommodations **once** during the State Testing window **and** may also apply for ACT Extended Time National Testing on any National Test Date.
   - Students whose disabilities require Special Testing (e.g., more than 50 percent additional time, alternate formats, or testing over multiple days) may request to test with accommodations **once** during the state testing window **and** may also apply for and test via ACT Special Testing during the months immediately before or after the statewide administration. To apply, students must submit a completed *Request for ACT Special Testing*.
2. <u>During 2012-2013.</u> Students who have been approved for ACT accommodations for the spring statewide administration and wish to retake the ACT with accommodations during the 2012-2013 academic year are eligible for a streamlined request process. These students will first need to determine which of the following options is appropriate to their disabilities:
   - ACT Extended Time National Testing for students who normally use up to 50 percent additional time and regular type or large type test booklets; **or**
   - ACT Special Testing for students who normally use more than 50 percent additional time, test over multiple days, or need alternate test formats (e.g., Braille, cassettes or DVDs, or a reader).

   These students will need to submit the appropriate 2012-2013 ACT request form with only Side 1 completed, along with a copy of their authorized accommodations letter from the statewide administration of the ACT. Requests for additional or different accommodations require a new request form completed in full with documentation to support the new accommodations.

ACT

IL

Application for Day 1 ACT-Approved Test Accommodations – Spring 2012
Prairie State Achievement Examination (PSAE)

The deadline for ACT to receive ACT-Approved Accommodations applications from your school is Friday, January 27, 2012.

This form is to be completed by a school official, such as a counselor, special education teacher, or principal, following the instructions provided in the *Procedures for Applying for ACT Test Accommodations – Spring 2012*.

A. STUDENT INFORMATION. (Please print or type.)

____
Student Name (Last, First, Middle Initial) Date of Birth (Mo/Day/Yr)

____
Student Street Address or PO Box City State Zip

______________________________ __ __ __ - __ __ __
Name of High School Where the Student Will Test ACT HS Code (required)
(This request must come in under the header sheet from the same school with the same ACT HS Code)

B. PREVIOUS APPROVAL OF THE SAME ACCOMMODATIONS ON THE ACT. Check either Yes or No to indicate whether this student has been approved previously for the same accommodations on the ACT and also has a current IEP or Section 504 Plan that supports the same accommodations that were previously approved.

- ☐ Yes If yes, complete all of Side 1 of this form and sign sections J and K. You may leave sections G, H, and I blank.
- ☐ No If no, both sides of this form must be completed and required documentation submitted.

C. DIAGNOSED DISABILITY. Check all that apply.

Learning Disability (01)
- ☐ (RD) Reading Disorder
- ☐ (DA) Mathematics Disorder
- ☐ (SL) Speech/Language Disorder*
- ☐ (DW) Writing Disorder/Written Expression

Physical/Sensory Disability (02)
- ☐ (DF) Hearing Impairment
- ☐ (PH) Motor Impairment*
- ☐ (VI) Visual Impairment*
- ☐ (TR) Tourette's Syndrome
- ☐ (EP) Epilepsy or Seizures

Psychological Disability (03)
- ☐ (AD) Attention Deficit Disorder/ADHD
- ☐ (AX) Anxiety Disorder*
- ☐ (BD) Emotional/Behavioral Disorder
- ☐ (AU) Autism Spectrum Disorder*
- ☐ (PD) Other Psychological/Cognitive Disability, including intellectual disability*
  FSIQ ______________

Other Disability (07)
- ☐ (HB) Confined to home*
- ☐ (OD) Other*

***Full documentation required**
(explain on side 2, G)

D. TEST FORMAT REQUESTED. Check only one. Alternate formats must be supported by diagnosis and IEP or 504 Plan. Examinees using reader's script must test individually. Readers may **not** read the tests to a group of examinees. For oral presentation, choose ONE of the following: DVDs, cassettes, or reader's script. **Note: If you do not check a box below, the student will automatically receive regular type (10-point).**

- ☐ (01) Regular Type (10-point)
- ☐ (02) Large Type (18-point)
- ☐ (03) Braille (printed copy included)
- ☐ (04) Cassettes w/ Regular Type
- ☐ (05) Cassettes w/ Large Type
- ☐ (06) Cassettes w/ Raised Line Drawings
- ☐ (07) Reader's Script w/ Regular Type
- ☐ (08) Reader's Script w/ Large Type
- ☐ (09) Reader's Script w/ Raised Line Drawings
- ☐ (19) DVDs w/ Regular Type
- ☐ (20) DVDs w/ Large Type
- ☐ (21) DVDs w/ Raised Line Drawings

E. TIME REQUESTED. Check only one. **ACT will assign a timing code (e.g., standard time, time-and-a-half, double time, triple time) based on the disability and approved test format.**

- ☐ Standard time - **large type only**
- ☐ Standard time on each test; authorization to test over multiple days
- ☐ Extended time on each test; authorization to test over multiple days
- ☐ Self-paced time-and-a-half, all tests on one day
- ☐ Extended time only on Writing Test (60 minutes)

F. OTHER ACCOMMODATIONS REQUESTED. Mark only if other accommodations are needed **in addition to** extended time or alternate formats (for example, authorization to use assistive technology), and enclose supporting documentation.

☐ Other (be specific):____

Student's Name (please print) City State

G. SPECIFIC DISORDER OR CONDITION. Complete only for those conditions marked with an asterisk (*) on side 1. Provide diagnostic, not narrative, information. If the diagnosis is not clearly stated, processing of the request will take longer and may require further information from the school before a decision can be made.

H. HISTORY OF DIAGNOSIS. If **FIRST** diagnosed *before* grade 9, complete only "age or grade of student" in section H-a., plus all information in section H-b. If first diagnosed *after* completing grade 8, all information requested in sections H-a. and H-b. must be completed. **COMPLETE DOCUMENTATION REQUIRED if FIRST diagnosed within last 3 years OR for visual, hearing, psychological, emotional, or physical disorders. (See "Guidelines for Documentation.")**

When and by whom student was:	H-a. **FIRST** diagnosed	H-b. recently re-confirmed **(within last 3 years)**
Date (month/year):		
Age or grade of student:		
Person making diagnosis:		
Name/team:		
Job title(s):		
Qualifications (degrees, specialization, certification):		

I. CURRENT IEP or 504 PLAN ON FILE AT SCHOOL. The IEP or 504 Plan must state the need for extended time, alternate formats, and/or any other accommodations requested on Side 1 due to the disability listed above. If plan has been in place **less than 3 years**, complete diagnostic documentation is required. **Note:** Only students who have an IEP or 504 Plan are eligible to apply for ACT-Approved Accommodations for PSAE Day 1.

1. Mark the appropriate box and attach the required copy (which must include student's name and effective dates).
   - ☐ IEP; attach a copy of the test accommodations/services page(s) from the current IEP.
   - ☐ 504 Plan; attach a copy of the test accommodations/services page(s) from the current 504 Plan.

2. Mark **ALL** school years for which the student has had an IEP or 504 Plan, **including year(s) before current school year**.
   ☐ 2011-2012 (grade 11) ☐ 2010-2011 (grade 10) ☐ 2009-2010 (grade 9) ☐ 2008-2009 (grade 8) ☐ Before grade 8

J. SCHOOL OFFICIAL'S SIGNATURE. *I affirm the student named on this form will test at this school, and I verify the information provided on this form and in the* ***attached IEP or 504 Plan and any other required documentation*** *is accurate, to the best of my knowledge and reflects the testing accommodations now provided in school.*

School Official's Signature (may not be a relative of the student) Print Official's Name and Title

School Official's E-mail Address

K. STUDENT/PARENT SIGNATURE. *I verify the information provided on this form is accurate to the best of my knowledge. I authorize the release to ACT of information related to this request by school officials, physicians, or others having such information, if requested. I understand that any documentation provided to ACT will remain with the application and will not become part of the student's permanent score record. If this request cannot be approved based on the information submitted, I understand the student may be required to test without the requested accommodations.*

Student's Signature (**required** if 18 or older) Parent/Legal Guardian Signature (**required** if student is under 18). ***Note:*** School official may sign for parent/legal guardian only if verbal acknowledgement has been obtained by phone. See *Procedures for Applying for ACT Test Accommodations.* Date

SUBMITTING THIS APPLICATION: Incomplete or unsigned forms will delay processing, which may result in the student having to test without accommodations. **Keep a photocopy for your files.** Early applications are encouraged. If ACT has questions about the information submitted, the Test Accommodations Coordinator will be contacted. The application must be submitted with a **completed** Test Accommodations Coordinator Header. Applications must be **received** at ACT by the appropriate deadline above and sent to:

ACT State Test Accommodations
301 ACT Drive
PO Box 4071
Iowa City, IA 52243-4071

ACT®

IL

Request for Day 1 State-Allowed Accommodations – Spring 2012
Prairie State Achievement Examination (PSAE)

The deadline for ACT to receive State-Allowed Accommodations requests from your school is **Friday, March 9, 2012**.

Important Note: Do NOT use this form to apply for ACT-Approved Accommodations. Complete this form ONLY for a student who does not meet ACT's eligibility requirements or whose application for ACT-Approved Accommodations has been denied by ACT. Scores earned with State-Allowed Accommodations will be used for PSAE purposes but will NOT be reported by ACT to colleges, scholarship agencies, or any other entities. PSAE Day 2 accommodations orders must be entered separately on PSAE TestSites Online. Contact the Student Assessment Division of the Illinois State Board of Education at 217/782-4823 for directions if a new student who needs accommodations enrolls in your school on or after the receipt deadline of March 9, 2012.

This form is to be completed by a school official, such as a counselor, special education teacher, or principal. Please review the *Procedures for Applying for ACT Test Accommodations – Spring 2012* for important information prior to completing this form.

A. STUDENT INFORMATION. (Please print clearly.)

____
Student Name (Last, First, Middle Initial) Date of Birth (Mo/Day/Yr)

____
Student Street Address or PO Box City State Zip

______________________________ ___ ___ ___ - ___ ___ ___
Name of High School Where the Student Will Test ACT High School Code (required)
(This application must come in under the header sheet from the same school with the same ACT HS Code)

B. REASON FOR REQUESTING STATE-ALLOWED ACCOMMODATIONS. Note: Only students who have an Individualized Education Program (IEP), Section 504 Plan, or who are Limited English Proficient (LEP) are eligible for State-Allowed Accommodations for PSAE Day 1. Check all that apply.

☐ (IEP) Individualized Education Program ☐ (504) Section 504 Plan ☐ (LEP) Limited English Proficiency

C. TEST FORMAT REQUESTED. Check only one. All test booklets, including large type, and all answer documents are printed in English. (Braille, if applicable, is normally an ACT-Approved Accommodation. If a student needs Braille in addition to other State-Allowed Accommodations, please call ACT at 800/553-6244, ext. 1788 before completing this request.) **Note: If you do not check a box below, the student will automatically receive regular type (10-point). ACT does not assign a timing code for students testing with State-Allowed Accommodations. The time allowed for each test is determined locally by appropriate staff at the school (e.g., IEP team).**

☐ (HB) Confined to Home

Printed Test Booklet (IEP, 504, LEP)
☐ (01) Regular Type (10-point)
☐ (02) Large Type (18-point)

Reader's Script (IEP, 504, LEP)
☐ (07) with Regular Type
☐ (08) with Large Type

Cassettes (IEP, 504, LEP)
☐ (04) with Regular Type
☐ (05) with Large Type

DVDs (Audio only) (IEP, 504, LEP)
☐ (DA) with Regular Type
☐ (DD) with Large Type

Spanish Video* (LEP)
☐ (DB) DVD with Regular Type
☐ (DE) DVD with Large Type
* More information about the Spanish format is posted on the PSAE website at www.isbe.net/assessment/SpDVD.htm

D. SCHOOL OFFICIAL'S SIGNATURE (required). *I affirm the student named on this form will test at this school. I have explained to the student and the student's parent/guardian that scores earned with State-Allowed Accommodations will be reported **ONLY** for PSAE purposes and will **not** be reported by ACT to colleges, scholarship agencies, or any other entities.*

______________________________ ______________________________
School Official's Signature (may not be a relative of the student) Print Official's Name and Title

E. STUDENT AND PARENT SIGNATURES (required). *I understand that scores earned with State-Allowed Accommodations will be reported **ONLY** for PSAE purposes and will **not** be reported by ACT to colleges, scholarship agencies, or any other entities. I understand that the student's notification of scores will be sent to the home high school in early fall.*

______________________ ______________________ __________
Student's Signature (**required** if 18 or older) Parent/Legal Guardian Signature (**required** if student is under 18) Date
Note: School official may sign for parent/legal guardian if verbal acknowledgement has been obtained by phone. See *Procedures for Applying for ACT Test Accommodations*.

SUBMITTING THE REQUEST: Incomplete or unsigned forms will not be processed. **Keep a photocopy for your files.** The request **must** be submitted with a **completed** Test Accommodations Coordinator Header. Requests must be **received** at ACT by the appropriate deadline above and sent to:
ACT State Test Accommodations
301 ACT Drive
PO Box 4071
Iowa City, IA 52243-4071

8/11 *01002012C* Rev 1

Appendix B

External Reviews of the Prairie State Achievement Examination

Page

External Review of the Prairie State Achievement Examination Reading and Writing Tests

by

Donna Ogle and Kenneth Hunter

The PSAE is a two-day, statewide academic examination that grade 11 public school students take each spring as required by state law. In February 2000—before ISBE made the decision to incorporate the ACT Assessment and WorkKeys *Reading for Information* into the PSAE—Illinois English teachers from across the state met to determine how well these tests cover the Illinois Learning Standards for reading and writing. They found that the ACT Assessment English Test thoroughly covers conventions (punctuation, grammar and usage, and sentence structure) and editing skills (strategy, organization, and style) and concluded that the ACT Assessment English Test when taken in conjunction with an ISBE-developed writing assessment matches the Illinois Learning Standards in State Goal 3, "*Write to communicate for a variety of purposes,*" extremely well. The English teachers also found there to be a good match between the ACT Assessment Reading Test and the Illinois Learning Standards for reading.

At the request of the Student Assessment Division of the Illinois State Board of Education (ISBE), we conducted an independent evaluation of the reading and writing portions of the PSAE, with an emphasis on the reading portion, to determine how well the PSAE reading and writing tests assess the Illinois Learning Standards for reading and writing. We also looked at all the Illinois Learning Standards for English Language Arts to determine how well the PSAE assessed the other language arts Standards. The analysis was conducted by the authors, Donna Ogle and Kenneth Hunter, educators who have direct experience with the secondary school reading curriculum, national and state standards for school reading programs, and the teaching and learning of reading at the high school level. Brief biographical summaries for both authors are attached to this report.

The central part of our review consisted of determining how well the PSAE tests assess the Illinois Learning Standards. In making that determination we also looked at two other tests that offer examples of what we believe to be improved ways of assessing reading comprehension. These two tests are the National Assessment of Educational Progress (NAEP) and the Program for International Student Assessment (PISA) reading assessments. The NAEP and PISA assessments are state-of-the-art assessments that are being used widely as reliable indicators of what is important for readers to be able to do in this new century. NAEP is a national measure designed to monitor the progress of American education. PISA was developed by the Organization for Cooperation and Economic Development (OCED), an intergovernmental organization of industrialized countries, as an international measure to assess the reading development of 15 year olds. The PISA framework was influenced by the NAEP design. We chose these two assessments to suggest possible directions for future testing because we are not aware of other standardized tests available for purchase that reflect this most current type of assessment.

To carry out our review and make pertinent comparisons, we created a matrix of the Illinois Learning Standards and Benchmarks for Language Arts and then mapped the PSAE components, NAEP, and PISA on that grid. Also as part of this review, we considered a number of questions that have been raised about the PSAE:

1. Students vary in their reading abilities. Are the passages sufficiently accessible so that students can demonstrate their comprehension and reading proficiency on the test?

2. Particular passages vary in their familiarity to students. Is the content of the passages related to students' prior knowledge? Do the texts include content that permits students to construct knowledge or are the passages so esoteric that they dissuade student engagement?

3. Is the content of passages related to the curriculum areas in which reading is important? Do passages map the kinds of reading students are asked to complete as part of their school experience?

4. How can students demonstrate their ability to summarize and respond interpretively, personally, and critically to texts they read?

Description of the Assessments

The PSAE Reading Test

The PSAE reading test is a combination of two assessments: the ACT Assessment Reading Test and WorkKeys *Reading for Information* assessment, both published by ACT and used nationally. ACT Assessment Reading is given on Day 1 of PSAE testing, and *Reading for Information* is given on Day 2. According to the ISBE *Teacher's Handbook* these assessments "test students' ability to read literary and informational texts with understanding and fluency."

The **ACT Assessment Reading Test** is one of the instruments in the ACT Assessment battery of tests, part of a curriculum-based assessment program. ACT Assessment Reading provides students with four passages to read and a total of 40 multiple-choice questions to answer (10 for each passage). The passages are selected from four areas: prose fiction, social science, humanities, and natural science.

Questions address the skills described in the ACT Standards for Transition®, which are statements of the skills and knowledge students in various score ranges are likely to have, and the Pathways for Transition®, which are a compilation of suggested activities to help students move from one score range to the next higher score range. These two resources can also be understood as a taxonomically arranged curriculum guide to the ACT Assessment. These materials are provided by ACT and are resources that teachers, principals, curriculum coordinators, and department chairs can put to effective use in classrooms.

The ACT Assessment Reading Test includes the following categories in which examinees demonstrate proficiency along a taxonomically staged score range:

- *Main Ideas*: Readers demonstrate proficiency along a continuum from the most basic task, "drawing simple conclusions about main points," to "identifying main ideas in…complex passages."

- *Significant Details:* In this category readers move through relatively "uncomplicated [to increasingly more] complicated" texts. They locate everything from "simple details" to finding and interpreting "subtly stated details [that]…support…idea or argument."
- *Sequence of Events:* ACT Assessment Reading asks readers to demonstrate ability in ordering sequence in both "uncomplicated and…complex passages."
- *Comparative Relationships:* The entry point of this area asks readers to "identify relationships between principal characters in uncomplicated passages." The difficulty range moves from identification to the highest point on the score range where readers are asked to "make comparisons, conclusions and generalizations in passages."
- *Cause-Effect Relationships:* Readers move from recognizing "clearly stated cause-effect relationships" in simple paragraphs to identifying "implied, subtle…cause-and-effect relationships" in even the most complicated selections."
- *Meaning of Words:* The degree of difficulty increases from using "context clues to understand basic figurative language" to a sophisticated skill level at which readers "determine the meanings of context-dependent words, phrases or statements" in any text.
- *Generalizations:* Here the reader is asked to "make simple generalizations" in most uncomplicated text settings to making "generalizations about people, ideas and situations…by synthesizing information from different portions…" of complex materials that may use "a range of literary devices."
- *Author's Voice and Method:* The most basic competency assessed in this area is the reader's ability to "recognize clear relationships between" the whole passage and its parts. Readers who demonstrate the greatest proficiency will be able to understand how those parts function "in relation to the whole…and then generalize about an author's… attitude or point of view."

The **WorkKeys *Reading for Information* assessment** is designed for a broader range of reading activities than the ACT Assessment and is described as representing informational reading needed in the workplace. The introduction to *WorkKeys: Helping to Build a Winning Workforce* explains that *Reading for Information* measures a person's skill in reading and using work-related information including instructions, policies, memos, bulletins, notices, letters, manuals and government regulations." *Reading for Information* is designed with passages at a range of reading levels, permitting students to demonstrate comprehension of real-world reading tasks.

Reading for Information comprises items grouped into levels of increasing difficulty. Examinees respond to 33 multiple-choice questions during the 45-minute test session. The passages have five levels of difficulty (Levels 3–7) designated by the test makers. Passages at level 3 are described as "short, uncomplicated texts which use elementary vocabulary such as basic company policies, procedures, and announcements. Questions focus on the main points of the materials and all information needed to answer the questions is stated clearly in the materials."

At Level 7, the highest level, the materials are more complex and more difficult than at the earlier levels, and the vocabulary is correspondingly more difficult. Jargon and technical terms whose definitions must be derived from context are included. The questions "require generalization beyond the stated situation, recognition of implied details, and recognition of the probable rationale behind policies and procedures."

The combination of ACT Assessment Reading and WorkKeys *Reading for Information* provides a richness of curriculum-connected and practical textual material for students to read. ACT Assessment reading passages reflect high school academic content and preview college work. *Reading for Information* extends the reading to include practical passages designed to reflect work-related situations and includes passages at a range of reading levels allowing students with less proficiency in reading ability to participate in demonstrating comprehension of reading tasks needed in the world of work.

The PSAE Writing Test

The PSAE assesses writing through the combination of the ACT Assessment English Test and the ISBE-developed writing test. The **ACT Assessment English Test** provides students the opportunity to demonstrate their proficiency in usage/mechanics and rhetorical skills as they apply rules in the context of five prose passages that students edit by selecting the best answer from multiple-choice test items.

The **ISBE-developed writing test** requires students to write an expository or persuasive essay in response to a single thematic or topical prompt. The scoring rubric has five features—focus, support, organization, conventions, and integration—and is used to assess students' ability to identify a topic and effectively communicate their views on that topic. The papers are written under timed conditions, so they are scored as first drafts with less emphasis on conventions than on the other features.

The two measures provide samples of a subset of writing skills. ACT Assessment English, with the emphasis on editing in context, provides a solid complement to the writing sample. It allows students the opportunity to show skill and knowledge in the conventions, while the writing sample provides them the opportunity to produce a complete document demonstrating their facility in composing and organizing text.

How the PSAE Assesses the Illinois Learning Standards for Reading

As required by **Standard 1B,** *Apply reading strategies to improve understanding and fluency,* students must be strategic readers to do well on the ACT Assessment Reading Test. However ACT Assessment Reading requires students to become strategic readers, but ACT Assessment Reading does not test whether students are aware of strategies that lead them to be successful in completing these tasks. Instead, students' use of strategies is inferred from their ability to respond correctly to test questions that address the categories described on pages 2 and 3 of this review, as can be seen in the following examples from an ACT Assessment Reading Test:

> *It can most reasonably be inferred that Anna and Emery attempt to deal with their cultural differences by:* (comparative relationships)

As it is used in line 82 the term Australopithecus *most nearly means:* (meaning of words)

According to the passage, if a mouse is reared in the dark during the first months of its life and later exposed to the light, it will never see normally because: (sequence of events/significant details)

Benchmark 1B 5a, "*Relate reading to prior knowledge and experience and make connections to related information*," is not addressed specifically in ACT Assessment Reading, although prior knowledge is certainly a contributing factor in students successfully navigating ACT Assessment Reading and *Reading for Information* passages: Knowledge of paleontology and biology would certainly be helpful in unpacking the meaning of the natural science selections; acquaintance with developmental psychology and political science would provide a platform from which students could more successfully access the social science passages; and a breadth of cultural knowledge would be of considerable use in moving successfully through the literature and humanities passages. Also, a sizable background vocabulary, considerable facility with etymology, and good word-attack skills are almost necessities for successful navigation of these texts.

Benchmark 1B 5b asks students to *"Analyze the defining characteristics and structures of a variety of complex literary genres and describe how genre affects the meaning and function of the texts."* ACT Assessment Reading offers selections from four areas—prose fiction, social science, humanities, and natural science—while *Reading for Information* provides selections from actual work-related materials. Students must have an understanding of genre and a working knowledge of the effect of text structure on writings to read these varied types of passages.

ACT Assessment Reading addresses this Benchmark through five of the categories described on pages 2 and 3 of this review: author's voice and method, significant details, main idea, comparative relationships, and cause and effect. Those categories are assessed in such items as the following:

The author does not mention volunteer work by name in this essay. Which of the following statements offers an explanation for this omission and is also supported by the essay? (author's voice and method)

The passage makes the claim that television news coverage is heavily influenced by Nielsen ratings because: (cause and effect)

Benchmark 1B 5c is "*Evaluate a variety of compositions for purpose, structure, content and details for use in school or at work*." This Benchmark addresses application of knowledge about text features and evaluation of author's effectiveness. We did not find this type of evaluative question on the ACT Assessment; neither does *Reading for Information* focus on evaluation of texts. Released samples from the NAEP reading assessment include a segment in which readers interact with official government documents through response to multiple-choice and constructed-response questions. In a 15-item question set students move back and forth through three documents to respond to questions asked. The final question of the set provides the opportunity for students to use all three documents—the W-2 form, the tax table and 1040EZ form—as they "complete (an) income tax return."

PISA offers a similar challenge for readers. In a more literary sample, readers are asked to interact with pro and con passages relating to two articles. Question sets require examinees to move fluidly between the two passages if they are to respond properly to the multiple-choice and constructed-response questions.

The areas most similar to the NAEP and PISA assessments on the two PSAE tests involve students being able to deal with items focused on the following categories: generalizations, main idea, significant details, comparative relationships, and author's voice and method.

Items such as the following support these categories as shown in these examples:

> *According to the passage, by reading her stories, many of the author's readers learned that:* (generalizations)
>
> *The main point of the passage is that:* (main idea)
>
> *The passage states that the ratio of brain weight to body weight in larger animals, compared to smaller animals, is:* (comparative relationships)
>
> *The author refers to Tom Sawyer (second paragraph, lines 11–23) to illustrate which of the following points:* (author's voice and method)

Benchmark 1B 5d states that students should be able to "*Read age-appropriate material with fluency and accuracy*." ACT Assessment Reading provides difficult—but age-appropriate—passages with extensive vocabulary from which students demonstrate their ability to make meaning through responses to multiple-choice questions. Although fluency and accuracy of reading are not tested directly, an indirect indication of fluency results from the timed nature of the tests and the amount of reading required: examinees who complete the test with high scores demonstrate both fluency and accuracy.

Items such as the following provide examples of questions that require accuracy in reading:

> *When the author asks "Why should nature have done that?" (line 74) which of the following questions is he really asking? (*sequence of events)
>
> *Which of the following statements most accurately expresses Fran's feelings when she hands her mother the letter from Linda Rose?* (cause and effect)
>
> *The author refers to Tom Sawyer in the second paragraph (lines 11–23) to illustrate which of the following points?* (author's voice and method)
>
> *In the fourth paragraph (lines 43–52), the author sets up a direct contrast between the image of the universe as a warehouse and:* (comparative relationships)

The ACT Assessment reading passages contain appropriately difficult words. The use of technical words, especially in such passages as "dinosaurs revised" and "participation in a modern democracy" (which also contains demanding nontechnical vocabulary), requires examinees to have both a rich vocabulary and a solid array of word-attack skills as required by **Standard 1A,** *"Apply word analysis and vocabulary skills to comprehend selections."*

Reading for Information provides passages that are arranged by difficulty. The *Reading for Information* levels are set from entry-level passages to much more demanding pieces. Examinees demonstrate both their fluency and accuracy through response to multiple-choice questions about the passage.

The intent of the Illinois Learning Standards for reading is that all students be able to read at grade level successfully. For example, the grade 3 Illinois Standards Achievement Test (ISAT) for reading does not contain grade 2 reading texts. However, it is clear that there are still great variations in students' reading abilities. The addition of *Reading for Information* with its varying levels of difficulty permits students with less-developed reading abilities to demonstrate their comprehension and fluency.

Standard 1C, *"Comprehend a broad range of reading materials,"* is addressed in the PSAE's use of ACT Assessment Reading and *Reading for Information*. Students are presented a wide array of textual materials representing a range of reading abilities. Their reading comprehension is addressed in the categories described on pages 2 and 3 of this review.

Benchmark 1C 5a requires that students be able to "*Use questions and predictions to guide reading across complex materials.*" Each question set for both ACT Assessment Reading and *Reading for Information* refers only to a single passage. While each passage is rich and complex, examinees do not have the opportunity to make use of questions and predictions across two or more texts at a time.

Benchmark 1C 5b states that students should be able to "*Analyze and defend an interpretation of text.*" ACT Assessment Reading offers multiple opportunities for students to meet this Benchmark. However, the ACT Assessment does not include students' defense of their own interpretations. They analyze and find evidence to support authors' statements and ACT-Assessment–given interpretations as shown in the following multiple-choice examples:

> *The author claims that the values he believes in are threatened by which of the following?* (generalizations)
>
> *The main point of the passage is that:* (main idea)
>
> *If the last paragraph were deleted, the passage would lose details about:* (sequence of events)
>
> *The author uses the description of the tax seminar in 1978 to make the point that some governmental issues are:* (author's voice and method)
>
> *The passage asserts that the octopus is more intelligent than:* (comparative relationships)
>
> *The author refers to the village of Faridpur as a phantom (line 27) because:* (meaning of words)

Benchmark 1C 5c states that students should be able to "*Critically evaluate information from multiple sources.*" ACT Assessment Reading and *Reading for Information* more than sufficiently meet a single source evaluation requirement, but they do not provide the opportunity to evaluate texts from multiple sources.

Benchmark 1C 5d states that students should be able to "*Summarize and make generalizations from content and relate them to the purpose of the material*." ACT Assessment Reading addresses this benchmark through two categories: generalizations and main idea. Sample items include the following:

> *It can be most reasonably inferred from the sixth paragraph (lines 60–80) that the Shaker belief system placed value on work that:* (generalizations)
>
> *One of the main points that the author seeks to make in the passage is that American citizens:* (main idea)

For students to actually demonstrate that they can summarize an assessment would require that they produce a written response. ACT Assessment Reading and *Reading for Information*, while asking students to identify main ideas and make generalizations through response to multiple-choice questions, do not allow them the opportunity for a constructed response or written summary. Students' ability to summarize accurately may, however, be inferred by their answers to these multiple-choice questions.

Benchmark 1C 5e states that students should be able to "*Evaluate how authors and illustrators use text and art across materials to express their ideas (e.g., complex dialogue, persuasive techniques)*." The ACT Assessment reading passages provide students the opportunity to interact with passages from a variety of areas. The prose fiction and humanities passages contain examples that address this Benchmark. The array of passages allows students to engage with different genres. The following examples include both text and test items:

The following is an excerpt from the prose fiction domain. The use of imagery "ghosts of all the long letters" is a key to selecting the appropriate response to a multiple-choice item.

> *I nodded and handed her the letter. It was short and businesslike, but I could see the ghosts of all the long letters she must have written and crumpled in the waste basket:*
>
> *Which of the following statements most accurately expresses Fran's feelings when she hands her mother the letter from Linda Rose:* ***Answer -*** *Fran knows how hard it must have been for Linda Rose to write the letter.*

The following is excerpted from a social science reading passage. This is a polemic focusing on the limits of democracy in a technological age. The author takes an ironic stance toward progress and provides rich and layered arguments to support his position. A number of items are used to assess student comprehension of the author's ideas:

> *The political orator of yesteryear has been replaced by a flickering image on the tube unlocking the secrets of the government universe in forty-five second licks. Gone forever are Lincoln-Douglas type debates… Newspapers take up the slack, but very little. Most of what one says to a local newspape… gets filtered through the mind of an inexperienced twenty-three year old journalism school graduate… Reporters focus on what sells papers or gets*

high Nielsen ratings; neither newspapers nor television stations intend to lose their primary value as entertainment.

Multiple questions are developed from this portion of the passage. They are listed below:

The author asserts that local newspaper reporters are often: ***Answer*** *- inexperienced and insufficiently educated.*

According to the passage, the news story under which of the following headlines would attract the greatest number of readers: ***Answer*** *- Senator Smith Claims 'I Never Made a Nickel On It.'*

The passage makes the claim that television news coverage is heavily influenced by Nielsen ratings because: ***Answer -*** *Television is an entertainment medium.*

Benchmark 1C 5f states that students should be able to "*Use tables, graphs and maps to challenge arguments, defend conclusions and persuade others.*" This reading task is not included in either ACT Assessment Reading or *Reading for Information*. While the PSAE does provide students the opportunity to work with tables, graphs, and maps in the ACT Assessment Science Reasoning, Mathematics, and ISBE-developed science and social science tests, ACT Assessment Reading and *Reading for Information* do not specifically address this Benchmark.

Clearly, the ability of students to read across texts and use graphic and visual information to build meaning are not assessed directly on the PSAE., nor is students' ability to summarize a text, to analyze and defend their own interpretation by showing their own work, or to compare texts on their own. Other formats, such as those on the more recently developed PISA and NAEP reading assessments, would be required for the test to measure these abilities. It is important to consider these other engagements as we think about what Illinois wants as part of its total assessment system, including local assessments, to ensure that the tests are assessing what our students should be capable of doing. Such skills become increasingly important as they reflect mature reading behaviors.

State Goal 2 requires that students be able to *"Read and understand literature representative of various societies, eras and ideas."* ACT Assessment reading passages are taken from the prose fiction, social science, humanities, and natural science arenas. The selections span eras and there is a bow to diversity, though the samples we reviewed were predominantly American pieces. However, the ACT Web site provides other sample passages that show a wider range of samples. The ACT Assessment provides *more than* sufficient representation of passages to meet the demands of this State Goal.

State Goal 3 requires that students be able to *"Write to communicate for a variety of purposes."* The writing ability of students is best measured through the ISBE-developed writing sample. In addition, ACT Assessment English assesses editing ability and awareness of English grammar and conventions. However, the PSAE does not include any extended writing in response to reading passage items, which would be useful in assessing the quality of the examinees' ideas about passages they have read more directly and fairly.

State Goal 4 requires that students be able to *"Listen and speak effectively in a variety of situations."* The requirements of standardized testing generally do not permit any use of assessments in which students demonstrate speaking and listening skills. ACT Assessment Reading, *Reading for Information*, NAEP, and PISA are paper-and-pencil tests in which student work in as much silence as possible. Alternative assessments, used at the local level can complement and support the teaching of this State Goal. For example, one district, Thornton High School District #205, has successfully developed and used such an assessment for more than 10 years. District #205's assessment instrument is modeled on the ISBE writing rubric and used to score student speech performance as the writing rubric is used to score student writing. As students in District #205 provide both a writing and speech sample, they have two opportunities to participate in the type of testing often called "authentic assessment." *The instrument is copyrighted by the district and, as such, does not appear in this review. Parties interested in using this assessment may contact Ms. Gwendolyn Lee, Assistant Superintendent for Curriculum in District #205.*

State Goal 5 requires that students be able to *"Use the language arts to acquire, assess, and communicate information."* ACT Assessment Reading and *Reading for Information* ask students to read and to actively engage with passages to make meaning from them. However, none of the items can assess the basic intent of Goal 5, which is that students independently use their reading and writing and search skills to engage in research and create their own reports of what they learn. The three standards require a more individual form of engagement and product creation. As is the case for State Goal 4, local assessment can do much to allow students to demonstrate proficiency in these areas.

The level at which the PSAE measures the skills and abilities needed to meet State Goal 5 is at the response level to given items. The assessments do allow students to demonstrate their abilities in acquisition and assessment of information through responses to multiple-choice questions in the categories described on pages 2 and 3 of this review as shown in the following samples:

> *In the context of the passage, what does the author mean when he states that "people…are scarcely worth mentioning" (lines 81–82)* (generalizations)
>
> *According to the first to paragraphs (lines 1–-16) researchers who study infant maturation want to find out:* Main Idea
>
> *Considering the information given in the first three paragraphs (lines 1–33), which of the following is the most accurate description of the author's girlhood and early adulthood?* (sequence of events)
>
> *In the fourth paragraph, the phrase "the triumph of hope over experience" (lines 57–-58) is an expression of the belief that:* (author's voice and method)
>
> *According to the information in the passage, if something were directly behind an octopus, would the octopus be capable of seeing it?* (significant details)

In the fourth paragraph (lines 43–-52), the author sets up a direct contrast between the image of the universe as a warehouse and: (comparative relationships)

The phrase visual field *(lines 33–-34) refers to:* (meaning of words)

Conclusions

The PSAE reading test must be seen as a unit. The Illinois Learning Standards and Benchmarks for reading cover a substantial range of knowledge and skills, not all of which can be easily assessed. Given the constraints of time and need for significance for the students taking the test, the use of ACT Assessment Reading and WorkKeys *Reading for Information* provides an acceptable basis for monitoring the progress of Illinois schools in meeting the Illinois Learning Standards for reading.

The inclusion of both ACT Assessment Reading and *Reading for Information* strengthens the test in three ways: It provides (1) a broad range of passage types, (2) a range of purposes for reading, and (3) passages with a range of reading difficulty. The inclusion of *Reading for Information* permits students the opportunity to show their comprehension and use of reading in real-world pieces. This is a real strength of the PSAE reading test and should be maintained. It should be noted, however, that there is a strong correlation (0.8) between ACT Assessment Reading and *Reading for Information* scores, indicating that student performance is consistent, regardless of the type of passage being presented to students.

It should also be noted that the PSAE reading test poses special difficulties for one particular group of students: those who are English-language learners (ELL). Specialized vocabulary is slow to develop in ELL students. Even many who have transferred out of bilingual programs lack the depth of vocabulary that permits success on the very short, unconnected passages that are generally used on standardized tests. The text and the assessment items are rich pieces and require facility with both language and culture, as examinees must interpret the meaning of passages and questions in context. Readers must bring an array of skills—in addition to direct translation—to the test, and ELL readers may be at a disadvantage in this arena. Teachers need to be aware of the difficulty that ELL students face and make sure that they are exposed in their regular classroom work to the kinds of texts and questions that appear on the ACT Assessment Reading Test and WorkKeys *Reading for Information.*

For the PSAE writing test, including both ACT Assessment English, which assesses editing grammar skills, and the ISBE-developed writing test, which allows students to demonstrate their ability to communicate their views in writing, thoroughly assesses State Goal 3.

Not all of the Illinois Learning Standards for English Language Arts are addressed by the PSAE *nor can they be appropriately addressed* in a two-day, timed, paper-and-pencil examination. So that these Standards are not neglected, the PSAE needs to be complemented by additional assessment pieces at the school and classroom level. Teachers need to be aware that the ISBE Standards Division has developed descriptors for all the Illinois Learning Standards for Language Arts and has collected high-quality examples of local assessments that are posted on the ISBE Web site.

Answering Our Questions

Students vary in their reading abilities. Are the passages sufficiently accessible so that students can demonstrate their comprehension and reading proficiency on the test? The PSAE reading test offers such accessibility to Illinois students through the combination of ACT Assessment Reading and *Reading for Information*. The passages that constitute the two assessments present materials that range from curriculum-oriented selections on ACT Assessment Reading to passages from the workplace on *Reading for Information*. Thus, the full assessment offers all students the opportunity to demonstrate proficiency in reading.

Particular passages vary in their familiarity to students. Is the content of the passages related to students' prior knowledge? Do the texts include content that permits students to construct knowledge or are the passages so esoteric that they dissuade student engagement? ACT Assessment Reading and *Reading for Information* both provide challenging passages. Prior knowledge, though not directly assessed by the ACT Assessment, is assuredly a factor in student performance. While none of the ACT Assessment reading passages that we reviewed were overly esoteric, those examinees with enhanced background information and well-developed read-to-learn skills would fare better in comprehending them. Superintendents, principals, curriculum directors, and department chairs would be well-served to review required curricular offerings along with enrichment opportunities for all students in the areas of prose fiction, social science, humanities, and natural science and in those areas that address the real world.

Is the content of passages related to the curriculum areas in which reading is important? Do passages map the kinds of reading students are asked to complete as part of their school experience? The four areas represented in ACT Assessment reading passages represent four of the core curriculum areas. It is our view that reading is not only important to these areas but of absolute necessity.

How can students demonstrate their ability to summarize and respond interpretively, personally, and critically to texts they read? ACT Assessment Reading and *Reading for Information* are multiple-choice formats. Students are asked to provide clear analysis of items related to passages as they are encouraged to make informed judgments in assessing the multiple-choice options. However, there is not the same opportunity to respond interpretively, personally, critically, and creatively that examinees are provided on other standardized assessments, such as NAEP, PISA, or the ISBE-developed reading ISATs. Those assessments provide the examinee a richer opportunity to make meaning from text through the inclusion of extended-response items and especially those in which multiple texts are involved. If these kinds of questions cannot be included on the PSAE, there should be an effort to promote their inclusion in local assessments.

Looking to the Future

The reading portion of the PSAE effectively allows students to demonstrate proficiency in meeting the Illinois Learning Standards. The pairing of ACT Assessment Reading and WorkKeys *Reading for Information* is a wise one. The college-oriented ACT Assessment Reading raises the bar in all Illinois classrooms and at the same time effects equity in that it requires all students to be exposed to high-quality reading experiences. The WorkKeys piece provides a needed complement and expands the types of reading passages to reflect more of the kinds of reading that students will encounter in their daily lives. This pairing of testing instruments establishes the PSAE reading test as a thorough assessment of students' reading proficiency in relation to the Illinois Learning Standards for reading.

While the PSAE is a solid assessment and ACT Assessment Reading and WorkKeys *Reading for Information* assess the Illinois Learning Standards, there are still some areas included in the state Benchmarks that are not addressed by the PSAE. These areas need to be addressed by local assessments. There is an increasing recognition that students need to read from multiple sources to develop their understandings of ideas and interpret events. Using graphic and visual information, reading and responding across multiple texts, critically evaluating texts, forming personal responses to texts, and reading and creating documents are essential in much of the learning students are asked to do. These are important skills for the twenty-first century, and all of these are Benchmarks included in the Illinois Learning Standards. Although inclusion of formats that measure these skills may not be feasible at the present time, when future test formats are considered, thought should be given to measuring these skills. To suggest possible directions for future testing, we included comparisons to the PISA and NAEP reading assessments in this review. We did not find any other standardized tests available for purchase that reflect this most current type of assessment. In any event, ISBE should emphasize the importance of these skills in local assessment programs and as essential elements of literacy.

Addendum to the External Review of the PSAE Reading Test

by

Donna Ogle and Kenneth Hunter

As expert reviewers of the PSAE Reading Test we are convinced that the Illinois Learning Standards (ILS) are adequately assessed through the two examinations that constitute the PSAE reading test. We want to clarify that Illinois's testing of high school students provides a sound measure of students' ability to meet the intent of the ILS. In the real world of student assessment, student proficiency on some of these reading outcomes and processes, while not directly measured on a group test, can be inferred from student performance. In particular, the PSAE reading test more than adequately assesses the Standards that pertain to reading: ILS 1A, 1B, 1C , 2A, and 2B.

ILS 1A requires students to *"Apply word analysis and vocabulary skills to comprehend selections."* As we state in our review, "The ACT Assessment reading passages contain appropriately difficult words. The use of technical words… requires examinees to have both a rich vocabulary and a solid array of word-attack skills." These same skills apply to the WorkKeys *Reading for Information* assessment, which includes specialized phrases, such as jargon and technical terms encountered in the workplace and in regulatory and legal documents.

ILS 1B requires students to *Apply reading strategies to improve understanding and fluency.* As we state in our review, "students must be strategic readers to do well on the ACT Assessment Reading Test." As we further make clear in our review, this Standard also applies to *Reading for Information*, which contains texts with a full range of difficulty, including instructions, policies, memos, bulletins, letters, manuals, government regulations, and legal documents.

ILS 1C requires students to *"Comprehend a broad range of reading materials."* As we state in the review, this Standard is addressed in both the ACT Assessment Reading Test and WorkKeys *Reading for Information*. Students are presented with an array of textual materials in both assessments. The WorkKeys assessment substantially broadens the variety of texts by its emphasis on nonacademic texts.

We understood the "reading across texts" concept in the Benchmarks that are included in this Standard to mean simultaneously responding to multiple passages, but a *reasonable and valid* interpretation of this Benchmark is that it refers to reading across a variety of texts. From this perspective, the PSAE reading test more than adequately meets this Standard. The ACT Assessment Reading Test and WorkKeys *Reading for Information* are two voices of literacy that offer a richness that certainly meets or exceeds the literacy requirements of ILS 1C.

Other Benchmarks in ILS 1C refer to the use of art, tables, graphs, and maps to express meaning in conjunction with text. The PSAE as a whole addresses these issues. The entire PSAE, which includes tests in science and social science as well as reading, writing, and mathematics, requires students to read, interpret, and evaluate tables, graphs, charts, maps, political cartoons, and other graphics. Although there is no federal requirement for students to be tested in these subjects, Illinois law requires that public school students take all the tests that constitute the PSAE. The Illinois 1994 AYP definition uses all subjects assessed in the grade 11 PSAE to generate a composite score that is used to determine AYP. (This composite score is for AYP use only; it is not reported to students or schools or contained in public reports.)

State Goal 2, which includes ILS 2A *(Understand how literary elements and techniques are used to convey meaning.)* and 2B *(Read and interpret a variety of literary works.)*, requires that students be able to

"Read and understand literature representative of various societies, eras and ideas." As we state in our review, **"**ACT Assessment reading passages are taken from the prose fiction, social science, humanities, and natural science arenas… The ACT Assessment provides *more than* sufficient representation of passages to meet the demands of this State Goal."

The PSAE reading test is a rich, challenging examination that raises the reading bar in every classroom in Illinois. The PSAE requires all students to demonstrate developed proficiency regarding the skills addressed in the Illinois Learning Standards. To meet the requirements of the PSAE, each classroom must become a focused space of enhanced reading opportunities. Classrooms must become places where each and every Illinois student is given the chance to thoughtfully and intelligently inter-act with a variety of texts from a wide array of reading voices. On the PSAE, each Illinois student is asked to apply such high-level skills as necessary to make meaning from a variety of rich and challenging passages representing a wide range of reading situations. These skills are important in the testing arena but find even greater application in the wider field of culture. The skills required by the Illinois Learning Standards, assessed through the PSAE, are those same skills essential to effective participation by Illinois students in their own lives and in the life of our democratic society. It is clear to this expert review team that the PSAE is a sound instrument that adequately assesses the Illinois Learning Standards and at the same time exerts a positive reading influence on each Illinois school and each Illinois classroom for each Illinois student.

External Review of the Prairie State Achievement Examination Mathematics Test

John A. Dossey
Sharon Soucy McCrone

The Prairie State Achievement Examination (PSAE) is the statewide academic examination that grade 11 public school students are required by state law to take each spring. This document reports an expert analysis of the contents and structure of samples of the two tests—the ACT Assessment Mathematics Test and WorkKeys *Applied Mathematics*—currently being used as the mathematics assessment of the PSAE. The analysis includes comparison of the PSAE tests with two other similar tests. The following tests were examined as part of this process:

- Mathematics Test, *ACT Assessment*, Form 58B, ACT, Inc., 1999.
- Mathematics Test, *ACT Assessment*, Form 58E, ACT, Inc., 1999.
- Applied Mathematics Test, *WorkKeys Assessment*, Form A07BB, ACT, Inc., 2001.
- Applied Mathematics Test, *WorkKeys Assessment*, Form C01BB, ACT, Inc., 2001.
- Mathematics Level IC Test, Form 3TBC2, The College Board, 1998.
- PISA Mathematical Literacy Test, OECD, 2000.

This analysis was made at the request of the Student Assessment Division of the Illinois State Board of Education (ISBE). In particular, the analysis was to accomplish the following objectives:

- Describe a model for analysis of the PSAE mathematics test,
- Identify and select one or more standardized mathematics tests for high school students in grades 10–12 that are generally recognized as having validity and credibility,
- Compare and evaluate the alignment of the PSAE and the other selected tests to the Illinois Learning Standards for mathematics for grade 11 students
- Compare and evaluate the quality of the PSAE mathematics test items and the PSAE mathematics tests as a whole with the other selected standardized tests for grade 11 students,
- Identify areas of strength and weakness in the PSAE relative to measurement of high school mathematics especially as related to the Illinois Learning Standards for mathematics for grade 11 students, and
- Present recommendations for improvement of the PSAE that would be feasible.

The present analysis was conducted from March to May 2002 by the authors, John Dossey and Sharon McCrone, mathematics educators who have direct experience with the secondary school mathematics curriculum, national and state standards for school mathematics, and the teaching and learning of mathematics at the high school level. Brief biographical summaries for both authors are attached to this report.

We began the analysis by first developing a framework based on a similar analysis made of the Illinois Standards Achievement (ISAT) tests for mathematics in 2001 (Dossey and Lindquist, 2001) and an analysis conducted by the U. S. Department of Education of the mathematics tests contained in the National Assessment of Educational Progress (NAEP), Third International Mathematics and Science Study, and the Program for International Student Assessment (Nohara and Goldstein, 2001). Once the framework was developed, each of us independently coded the items of the tests included in the study for each of the variables of the framework. We then met to discuss our individual analyses and to develop the final codes that serve as the basis for our discussion of the tests. Finally, we jointly developed the present report detailing our analysis and findings.

Description of the Prairie State Achievement Examination

Information in this section is from the ISBE Web site (http://www.isbe.net/) and was downloaded on March 24, 2002. On that date, the site indicated that the information was last updated on March 12, 2002. Some material has been deleted, but the essence has been retained to provide an ISBE-developed definition of the nature and goals of the PSAE.

> The PSAE includes three components: (1) ISBE-developed writing, science, and social science assessments; (2) the ACT Assessment, which includes reading, English, mathematics, and science reasoning; and (3) two WorkKeys assessments (*Reading for Information* and A*pplied Mathematics*). Thus, the mathematics section of the PSAE has two components: the ACT Mathematics Assessment, taken on Day 1, and WorkKeys *Applied Mathematics*, taken on Day 2. The scores of these two examinations are combined to produce the PSAE mathematics score.
>
> The PSAE has two purposes: (1) to measure student progress toward meeting the Illinois Learning Standards for school accountability and (2) to recognize the achievement of individual students who receive a Prairie State Achievement Award for excellent performance.
>
> Illinois gives the PSAE because it measures student progress toward meeting the Standards and provides additional benefits to students, including ACT Assessment and WorkKeys scores. As originally passed in 1996, the PSAE legislation would have required ISAT to continue at grade 10 (for reading, writing, and mathematics) and grade 11 (for science and social science). In addition, the PSAE was to assess reading, writing, mathematics, science, and social science at grade 12. Before this statewide high school testing program could be implemented, ISBE worked with legislators to make changes so that high school testing would be reasonable for schools. The current legislation, passed in 1999, eliminated ISAT at grades 10 and 11 and established the PSAE as the only mandated statewide academic assessment beyond grade 8. The PSAE was administered for the first time in spring 2001. ISBE has contracted to use the ACT Assessment and two WorkKeys assessments through 2005.
>
> Students are allowed to use certain types of calculators on the mathematics portion, but <u>not</u> on tests for other subjects. Types of calculators that may be used for the respective mathematics tests are described in *Preparing for the ACT Assessment 2001–2002* and on page 52 of the PSAE student test-preparation booklet, *Overview and Preparation Guide for PSAE Day 2*. In addition, details about calculators are available on the ACT Web site at <u>www.act.org</u>. Students are responsible for supplying their own calculators; schools may, if they wish, lend calculators to students who need to borrow one.
>
> A <u>formula sheet</u> is provided as part of the test booklet for the WorkKeys *Applied Mathematics* assessment. However, students are <u>not</u> allowed to use a formula sheet for the ACT Assessment

Mathematics Test. Students need to know basic formulas and perform basic computational skills to solve problems on the ACT Assessment Mathematics Test, but do not need to know complex formulas or perform extensive computation.

Students receive a PSAE scale score and performance-level designation for each of the five subjects assessed by the PSAE. In addition, the PSAE also generates the following scores from the ACT Assessment and two WorkKeys assessments:

- An ACT Assessment Composite Score
- ACT Assessment Scores [four tests in caps and seven subtests in italics]

  ENGLISH – *Usage/Mechanics* and *Rhetorical Skills*

  MATHEMATICS – *Pre-Algebra/Elementary Algebra*, *Intermediate Algebra/Coordinate Geometry,* and *Plane Geometry/Trigonometry*

  READING – *Social Studies/Sciences* and *Arts/Literature*

  SCIENCE REASONING

- WorkKeys Test Scores [2 test scores in caps]

  READING FOR INFORMATION

  APPLIED MATHEMATICS

The Tests

The PSAE comprises two separate mathematics tests, the ACT Assessment Mathematics Test and WorkKeys *Applied Mathematics* test. Scores from these two tests are combined to give each Illinois student a PSAE scale score and a performance level in mathematics. The individual scores from the ACT Assessment and WorkKeys *Applied Mathematics* and the subtests of the ACT Assessment are reported to students as well. Before ISBE adopted the PSAE—at the time that the ISAT was the mandated statewide test for public high school students—Illinois students took an examination that was developed by ISBE in collaboration with its test-development contractor and Illinois teachers. This is not the case with the PSAE. Although Illinois teachers may apply to become item writers for the ACT Assessment or apply to participate as item writers and reviewers for the WorkKeys assessments, ISBE has made extensive materials, including released ACT Assessment test forms and released WorkKeys and ISBE-developed test items, available for teachers and schools in both print and electronic forms to help them understand the tests that constitute the PSAE and what they need to do to familiarize their students with the requirements of these tests. In what follows, we give a brief overview of both mathematics tests in the PSAE. In addition, we provide a description and review of two other grade 11 tests, the SAT II, Level 1C examination and the PISA mathematics literacy assessment, which we reviewed and compared to the PSAE tests.

The **ACT Assessment Mathematics Test** is a 60-item, multiple-choice test with 5 response options for each question. It has a 60-minute time limit. The test is written to assess the mathematical concepts and skills that students have typically acquired prior to grade 12. The test design assumes a command of basic definitions, algorithms, and formulas. Students are expected to know basic formulas and mathematical relationships. When a formula beyond the basics for area and volume is required, it is provided in the item. Students are allowed to use a calculator while taking the test. The calculator must be from an ACT-approved list of calculators. This list includes common scientific and graphing calculators, but does not allow the use of calculators with QWERTY-keyboards.

The ACT Assessment Mathematics Test includes a wide range of items that address general mathematics knowledge and skills, direct applications of these skills, understanding of concepts, and an integration of conceptual understanding and procedural knowledge. In addition, the test is designed to provide a basis for an overall score as well as subscores in pre-algebra/elementary algebra (24 items), intermediate algebra/coordinate geometry (18 items), and plane geometry/trigonometry (18 items). The framework for the test suggests: pre-algebra (23 percent of test, 14 items); elementary algebra (17 percent, 10 items); intermediate algebra (15 percent, 9 items); coordinate geometry (15 percent, 9 items); plane geometry (23 percent, 14 items); and trigonometry (7 percent, 4 items) (ACT, 2001).

The **WorkKeys *Applied Mathematics* Test** is a 33-item, multiple-choice test with 5 response options for each question. It has a 45-minute time limit. The test is written for a multitude of purposes, including job-profiling, personnel assessments, instruction support needs, and reporting for businesses and educational institutions. The test provides students with a formula sheet containing basic measurement conversions (including linear and nonlinear measurements, electricity, and temperature) and common area and volume formulas. Students are allowed to use any calculator on the ACT list in taking the test.

The *Applied Mathematics* test is designed to measure a person’s skill in using mathematical reasoning to solve work-related problems. Test takers set up and solve problems similar to those that would occur in a workplace. Scores represent five levels of achievement from a low of <3 to a high of 7, that correspond to command of a variety of mathematics skills. For example, an examinee at Level 5 can work appropriately with common conversions of units, calculate in a several-step problem situation, calculate percentages of increase and decrease, and determine what information is required and what strategy is valid to solve a problem. An examinee at Level 7 can calculate using several steps involving logic, calculate areas in problems requiring the manipulation of several subareas, solve problems with more than one unknown, handle rates of change in nonlinear settings, and apply basic statistical concepts (ACT, 2000).

The **SAT II, Level IC Mathematics Test** is a 50-item, multiple-choice test with 5 response options for each question. It has a 60-minute time limit. The test is written as a placement test for colleges and universities for use in bringing secondary school students into their programs at the appropriate level. The test provides students with a formula sheet containing basic measurement conversions and common area and volume formulas. Students are allowed to use any calculator on a specified list of calculators in completing the items on the test. This list is similar to the ACT list and also excludes the use of calculators with a QWERTY keyboard.

The SAT II, Level IC test is built on the expectation that the students taking it will have had at least three years of college-preparatory mathematics, including two years of algebra and one year of geometry.

The test is designed to help place students who have completed such a sequence into appropriate college courses. As such, its composition is similar to that of the ACT Assessment Mathematics Test. The composition of test items by area of mathematics is essentially: algebra, 30 percent; plane geometry, 20 percent; coordinate geometry, 12 percent; three-dimensional geometry, 6 percent; trigonometry, 8 percent; functions, 12 percent; statistics, 6 percent; and miscellaneous, 6 percent. The latter category contains items that address number theory, logical reasoning, and similar topics found in almost all mathematics programs.

The **PISA Mathematical Literacy Test** is a 32-question, mixed-item format test. It has a 60-minute time limit. The test was developed as part of an international assessment of 15-year-old students (U. S. Department of Education, 2001). As such, it focuses on students' ability to apply mathematical principles and thinking in a wide variety of situations. The test was designed to assess the mathematical literacy level of countries' 15-year-old populations as a proxy for their future capacity to manage change in a technological world. Students were allowed to use any calculator that they normally used during instruction in taking this examination.

The PISA Mathematical Literacy Test is constructed to measure students' command of the processes and content of mathematics in context. The processes involve students' developed capabilities in mathematical thinking, mathematical argumentation, modeling, problem posing and solving, representation, symbols and formalism, communication, and use of aids and tools. The items are divided into levels of competence: reproduction, definitions, and computations; connections and integration for problem solving; and mathematization. Mathematization measures a students' ability to consider a situation, abstract out the mathematics, generalize it if necessary, build a model, solve the problem, and reflect on the solution. Several of these steps are built around creative work on the part of the individual student.

All of tests reviewed in this study are built on sound psychometric grounds and have been examined from both a reliability and validity standpoint. While they were developed to serve different purposes, they are sound tests. We selected the SAT II, Level IC and PISA tests to compare and contrast with the ACT Assessment Mathematics Test and WorkKeys *Applied Mathematics* for two reasons. First, these tests bear a similarity to the mathematics portions of the PSAE. ACT Assessment Mathematics and the SAT II Level IC are mathematics tests that purport to have as a base prerequisite an understanding level of Algebra II. The WorkKeys mathematics and PISA tests purport to address understanding and applying mathematics in real-world contexts. The second factor for our choices was that the SAT II series of tests and the PISA instrument were developed in the same time frame as the PSAE components and are widely known and recognized.

The Analysis Framework: the Variables

Several studies have been made that compare the content of extant assessments relative to content and cognitive frameworks related to the programs for which the assessments serve (Dossey, 1996; Dossey, Peak & Nelson, 1997; Gandal & Dossey, 1997; McLaughlin, Dossey & Stancavage, 1997; Burrill, Paulson, Dossey & Webb, 1998; Nohara and Goldstein, 2001; and Dossey & Lindquist, 2001). Relying on the general framework of several of these studies and the mathematics portion of the Illinois Learning Standard*s* (ISBE, 1997), we decided to code the tests using the following variables : the content tested by an item, the cognitive demand of an item, the presence of a real-world context in an item, whether an item requires computations, whether a calculator would have been of assistance in completing an item, the number of steps a student probably would have taken in completing an item, and whether an item involved a representation (graph, drawing, data table, or other auxiliary formatted information) that a student had to decode in addition to the written statement of the problem. Each of these variables is described in greater detail in the following subsections.

Content

The content categories used for the analysis were as defined in the *Item and Test Specifications* (ISBE, 1998). Each item on the tests was coded relative to our judgment of which single content category best described the mathematics content being assessed by the item. These categories are as follows:

1. Estimation/Number Sense/Computation. Includes items that may require students to demonstrate an understanding of numbers and their representations, estimate and perform number operations involving addition, subtraction, multiplication, division, percentages, fractions, ratios and proportions of rational and irrational numbers, as appropriate to the level of schooling. (Illinois Learning Standards 6A, 6B, 6C, 6D, 8C).
2. Algebraic Patterns and Variables–. Includes items that may require students to identify, describe, and extend geometric and numeric patterns and to construct and solve problems using variables, as appropriate to the level of schooling. (Illinois Learning Standards 8A, 8D)
3. Algebraic Relationships/Representations. Includes items that may require students to represent and interpret algebraic concepts with words, diagrams, tables, function notation, number lines, coordinate graphs, equations and inequalities, as appropriate to the level of schooling. (Illinois Learning Standard 8B)
4. Geometric Concepts. Included items that may require students to identify and describe points, lines, angles, two- and three-dimensional shapes and their properties (including the Pythagorean Theorem). May also include topics involving symmetry, parallel and perpendicular lines, and number of sides, faces, or vertices, as appropriate to the level of schooling. (Illinois Learning Standard 9A)
5. Geometric Relationships. Includes items that may require students to sort, classify, compare and contrast geometric figures. They may include properties such as similarity and congruency, as appropriate to the level of schooling. (Illinois Learning Standards 9B, 9D)
6. Measurement. Includes items that may require students to estimate, measure, compare and convert (within measurement systems) quantities using appropriate units and acceptable levels of accuracy. May include items that involve computing area, surface area, and volume, as appropriate to the level of schooling. (Illinois Learning Standards 7A, 7B, 7C)
7. Data Organization and Analysis. Includes items that may require students to create, analyze, display, and interpret data using a variety of graphs. May include items such as pictures, tallies, tables, charts, bar graphs, and Venn diagrams and the computation of mean, median, mode, and range for a set of data, as appropriate to the level of schooling. (Illinois Learning Standards 10A, 10B)
8. Probability. Includes items that may require students to determine, describe, and apply the probability of an event and to use fundamental counting principles such as permutations and combinations or simple and complex events, as appropriate to the level of schooling. (Illinois Learning Standard 10C)

These eight categories were maintained throughout the coding process. By combining categories 2 and 3, 4 and 5, and 7 and 8, one can collapse these eight categories into the five learning areas of number, measurement, algebra, geometry, and data analysis and probability that are used in the Illinois Learning

Standards (ISBE, 1997), the NCTM *Principles and Standards for School Mathematics* (NCTM, 2000), and the National Assessment of Educational Progress (NAGB, 1994).

Cognitive Demand

Each test item was classified with respect to cognitive complexity: the cognitive demand an item might place on grade 11 students currently enrolled in an Algebra II course. The value we assigned was a professional determination of the demand relative to students' potential opportunity to learn the content required and what they might reasonably have been expected to do with that content in their learning of it. We defined four categories—routine, nonroutine, simple, and complex—which constitute the variable of cognitive demand. Any given item can contain information that students have directly studied (routine) or that they most probably have not seen directly as part of their learning (nonroutine). The task presented can be somewhat direct and similar to actions the student has practiced a number of times (simple) or can be more demanding in the processes the student is asked to perform (complex). Complex items are those requiring analysis, synthesis, and evaluation and are items that the students probably had little or no practice with as part of their mathematics learning experiences.

These four categories define a 2 × 2 model for cognitive demand illustrated in Table 1. The four levels for cognitive demand are simple-routine, complex-routine, simple-nonroutine, and complex-nonroutine. They form a hierarchy of knowing and doing mathematics, at least as related to students' opportunity to learn and acquire familiarity through investigation and practice. This model is similar to that proposed for the framework for NAEP 2005 (NAGB, 2001).

Table 1: Cognitive demand categories and their weights

	Routine	Nonroutine
Simple	1.0	1.6
Complex	1.4	2.0

The weights shown in Table 1 reflect our view of the relative demand such items place on the learner and were used to analyze the relative overall demand placed by examinations on students. The cognitive demand of an item is not a function of the format in which it is presented (multiple-choice, short constructed response, extended constructed response), as any particular format can be found in each of the demand categories.

Item Format

One of the critical variables of concern in this analysis is the nature of the response format created by the types of item. Items on a test could be multiple-choice, simple constructed response, or extended constructed response. A simple constructed response item asks only for a computation or an identification type of response and is scored on a right-wrong basis or, at most, a 0-1-2 rubric. An extended constructed response item calls on students to provide some rationale and some form of communication about their work on the problem. Extended constructed response items could be graded with a 0-1-2 through 0-1-2-3-4 rubric.

Context

The variable of context refers to whether an item is posed in a real-world setting or is given as a naked mathematics item. The context of an item is important for a number of reasons. First, context can make items either difficult or easy. In some cases, an unfamiliar context can lead a student to avoid an item, even when the mathematics involved is familiar and rather easy. In other cases, the context serves as a motivator for students, particularly if the context is familiar to the student. Context can increase the reading load for an item and create extra representational translations from text to symbols or from diagrams to symbols to graphs. However, one goal of a mathematics curriculum is to educate students to function in context-rich situations. Students need to be able to translate from real-world settings to mathematics settings, solve the problem, and then translate the answer back into the real-world setting. Items were coded as a 0 if they had no real-world context or only a hint of context, such as using the term "rubber ball" rather than the more mathematical term "sphere." Items were coded as a 1 if they were set in a real-world context or referred to physical objects different from mathematical objects, such as a barn roof or a map.

Computation

Items were also examined to see if there was any calculation involved in finding the solution to the problem posed. If a calculation of any type was called for in the solution of a problem, it was coded as a 1 on this variable. If no calculations were needed, then the item was coded as a 0. This variable gives an indication of the number and operation load in an examination, which is important because even though an examination may be balanced in terms of number sense, measurement, geometry, algebra, and data and probability, a high value on the calculation variable indicates that the assessment has a high reliance on students' knowledge of number and operation, one far beyond what is indicated by the percentage of items coded as number sense and computation. While it is not always possible to ascertain the way in which students might work a problem, our best guesses served as the guide for this coding.

Calculator Usage

The variable "calculator use" was added to the analysis to measure the effect calculator usage might have on student performance. As all examinations allowed calculators, an item was scored as a 1 on this variable if it involved an operation with numbers that called for more knowledge than the basic facts associated with the four whole number operations. That is, the item was scored a 1 if it included such forms as fractions, decimals, and integers, or if it included calculations with whole numbers beyond those associated with the basic facts. If the problem could be solved with no calculations or only involved a simple, basic-fact calculation with whole numbers, then it was scored as a 0. Some items were also scored as a 0 if they involved simple calculations with square roots or fractions in which the decimal approximation of the root or fraction was not helpful in determining the correct answer. While it is not always possible to ascertain the way in which students might work a problem, our best guess served as the guide for this coding.

Multistep Thinking

Items were coded as involving either one step or two-or-more steps depending on our best determination of the way grade 11 students might attempt to solve a problem. If a given item involved adding several numbers, such as a typical column addition problem, it was scored as a 1, as it basically involved one string of adding. In the case of finding the average of a group of numbers, the problem was coded as a 2, for in this case the students first would have to add to get the total and then divide to get the average of the numbers. In general, a 1 indicates a problem in which the student has merely to select an operation and perform it. A 2 indicates a problem in which one operation first has to be accomplished before the next portion of the problem can be attempted. As in the previous descriptions of variables for item analysis, the scoring for multistep thinking involved a value judgment on our part.

Representation

In addition to the seven variables described in the preceding subsections, items were classified in one additional manner. They were coded as a 1 for including a representation if the students had to interpret a graph, chart, table, drawing, and think about or use a manipulative aide (such as a spinner or dice) for completing the problem. Such items were defined as involving a representation. Items were coded as a 0 if they involved no representations other than a verbal or symbolic representation, such as is usually found in written mathematics. If an item was coded as a 1, then a second coding was performed to indicate the type of representation involved. The codes for this portion of the analysis were used to indicate that the item involved the following types of representations:

1. Geometric figure or diagram
2. Algebraic graph on a coordinate axis
3. Number line
4. Data table, a matrix, or a structured listing of data or numbers
5. Statistical graph of some type
6. Some form of probability representation, such as a spinner or dice
7. Scale drawing or similar figure interpreted by a scale
8. Sketch with measurements for area or volume problems
9. Representation of terms in an algebraic or geometric pattern.
10. Photograph

We met after individually coding the items and reconciled our judgments, concluding with the data reported in the following section.

The Findings

This analysis of the PSAE mathematics tests, the SAT II Level IC mathematics examination, and the PISA mathematical literacy test found a good deal of differences among the tests. Further, analysis of different forms of the same test found a degree of variation within forms of a given test. In the following sections, the data from each of the variables are depicted, then analyzed and commented upon.

Content

Any analysis of content must be based in what are appropriate emphasis levels for the five content areas highlighted by the Illinois Learning Standards for mathematics: number sense, estimation and measurement, algebra and analytic methods, geometry, and data and probability. One accepted basis for such a comparison are the emphasis percentages given by NAEP, the Nation’s Report Card (NAGB, 1994, 2001) for its grade 12 assessments shown in Table 2.

Table 2: Recommended percentages for emphasis on grade 12 NAEP – 1996, 2000, and 2005

<table>
<tr><th>Content Area</th><th>1996 and 2000</th><th>2005</th></tr>
<tr><td>Number sense</td><td>20%</td><td>10%</td></tr>
<tr><td>Measurement</td><td>15%</td><td rowspan="2">30%[a]</td></tr>
<tr><td>Geometry</td><td>20%</td></tr>
<tr><td>Data analysis and probability</td><td>20%</td><td>25%</td></tr>
<tr><td>Algebra</td><td>25%</td><td>35%</td></tr>
</table>

[a] The recommendation is that in 2005 geometry and measurement combined make up 30 percent of the questions.

This analysis shows a marked decrease in emphasis on number sense at grade 12, a slight decrease in emphasis on geometry and measurement, a slight increase in data and probability, and a marked increase on algebra. These recommendations also parallel the weights suggested by the NCTM's *Principles and Standards for School Mathematics* (NCTM, 2000).

As indicated in Table 3, both forms of the ACT Assessment have a high percentage of items in the area of algebra, which compares well with the SAT II and is not far from the recommended weighting given in Table 2. The lower number of items in number and operations of the ACT Assessment and the SAT II corresponds with NCTM recommendations that basic skills be maintained throughout high school although the focus of learning need not be in this area (NCTM, 2000). Both forms of the ACT Assessment we examined have only about 20 percent of their items in the area of geometry. Although some of the measurement items may be considered to contain geometric content, even the sum of these two categories leaves the percent of geometry items below that of the SAT II, which is more balanced between algebra (42 percent of items) and geometry (38 percent of items).

Table 3: Number and Percent of Items Relative to the Illinois Learning Standards.

	ACT Form 58B		ACT Form 58E		WorkKeys A07BB		WorkKeys C10BB		SAT II- 1C Form 3TBC2		PISA Math Literacy	
	#	%	#	%	#	%	#	%	#	%	#	%
NUMBER	8	13	10	17	22	67	20	61	4	8	3	9
MEASUREMENT	8	13	7	12	9	27	10	30	2	4	4	13
ALGEBRA	(27)	(45)	(24)	(40)	(0)	(0)	(0)	(0)	(21)	(42)	(8)	(25)
Patterns & Variables	13	22	12	20	0	0	0	0	13	26	3	9
Relations/ Representation	14	23	12	20	0	0	0	0	8	16	5	16
GEOMETRY	(11)	(19)	(14)	(21)	(0)	(0)	(0)	(0)	(19)	(38)	(7)	(22)
Concepts	7	12	3	5	0	0	0	0	12	24	1	3
Relations	4	7	11	16	0	0	0	0	7	14	6	19
DATA/CHANCE	(6)	(10)	(5)	(8)	(2)	(6)	(3)	(9)	(4)	(6)	(10)	(31)
Data Analysis	4	7	3	5	2	6	3	9	3	4	10	31
Probability	2	3	2	3	0	0	0	0	1	2	0	0

With growing emphasis on data analysis in education as well as in the workplace and everyday life, it is surprising that all tests except the PISA assessment contain very few items in the areas of data and probability. Even the WorkKeys test contains very few items in this area.

In comparison with the PISA assessment, the other five tests are not as balanced across the five content areas. The ACT is comparable to the SAT II in all areas except geometry, as described previously in this subsection. The WorkKeys tests, however, are heavily laden with number and operations items as well as measurement items. One of the stated goals for WorkKeys *Applied Mathematics* is to test students' ability to solve mathematics problems from the workplace. Considering only the data in Table 3, it appears that *Applied Mathematics* assesses mainly basic number skills. Based on the Illinois Learning Standards, it would appear that ISBE would want to be assured that students are able to employ their basic number skills to solve a broad range of uses of mathematics across measurement, geometry, data analysis, chance, and algebra, as well as rather straightforward applications of basic number operations.

Cognitive Demand

The ACT Assessment and WorkKeys *Applied Mathematics* are comparable in their cognitive demand on all levels. The SAT II and the PISA tests appear to differ significantly from the PSAE tests and from each other in the number of items coded as either simple-routine or complex-routine. On the one hand, the PISA test is less cognitively demanding than the other tests, while the SAT II appears to be more demanding.

Table 4: Number and Percent of Items by Cognitive Demand Categories

		Number of Items		Percent of Items	
		Routine	Nonroutine	Routine	Nonroutine
ACT Form 58B	Simple	23	16	38	27
	Complex	16	5	27	8
ACT Form 58E	Simple	19	20	32	33
	Complex	11	10	18	17
WorkKeys A07BB	Simple	14	7	42	21
	Complex	7	5	21	15
WorkKeys C01BB	Simple	14	10	42	30
	Complex	4	5	12	15
SAT II-1C From3TBC2	Simple	7	16	14	32
	Complex	19	8	38	16
PISA Mathematical Literacy	Simple	20	5	63	16
	Complex	4	3	13	9

Part of this difference results from the fact that the PISA test is an assessment of mathematical literacy, not achievement. It is focused on what students can do with their mathematical knowledge when confronted with a problem from the real-world. While similar in nature to the WorkKeys test in focusing on nonschool/noncurriculum items, the PISA assessment items tend to reach more into unique areas involving environmental issues, barn construction, and common sense interpretation of quantitative relationships, while the WorkKeys items focus on specific applications that might be found in the workplace.

Item Format

Table 5 presents the results of an analysis of the items found on the various tests that were included in this study. The items were categorized in terms of multiple-choice, short answer, and extended responses as defined earlier. The comparisons showed a great deal of similarity in the ACT, WorkKeys, and SAT II examinations. These examinations were entirely composed of multiple-choice items. The PISA test, on the other hand, presented students with a balanced set of items, similar to what is found on NAEP on which the balance of items at the grade 12 level in the recent past has been approximately 60 percent multiple-choice, 35 percent short answer, and 5 percent extended responses (Braswell et al., 2000).

Table 5: Number and Percent of Items by Response Formats

	Multiple-Choice		Short Answer		Extended Response	
	Number	Percent	Number	Percent	Number	Percent
ACT Form 58B	60	100	0	0	0	0
ACT Form 58E	60	100	0	0	0	0
WorkKeys A07BB	33	100	0	0	0	0
WorkKeys C01BB	33	100	0	0	0	0
SAT II Level 1C	50	100	0	0	0	0
PISA Math. Lit.	11	34	15	47	6	19

The analysis of the balance of items in the PSAE indicates that students were expected to do little in terms of meeting the objectives that are stated in ISBE's "Applications of Learning" in terms of solving problems, communicating, using technology, and making connections. These cognitive process objectives, which proceed to ISBE's statement of specific learning standards in mathematics, reflect the cognitive processes and skills students are expected to develop and be able to use as a result of their study of mathematics. When students are expected to produce extended responses to items on an examination, they are driven to make connections, to reason and structure communications, and to think through and actually solve problems, not just select answers. Such items are also less susceptible to test-taking strategies than are multiple-choice items. As such, only the PISA assessment comes close to matching the NAEP criteria or the balance of items that one would expect from a test that measures a wide range of cognitive objectives. If the state of Illinois is serious about students solving problems and communicating in mathematics, it must place extended-response items requiring both short answers and extended answers on its PSAE.

Context

The next category we investigated was the amount of context that appeared in the problems presented. The Nohara (2001) analysis of TIMSS-R, NAEP, and PISA at the grade 8 level indicates that TIMSS-R and NAEP both had context present in about 45 percent of their items, while context was a part of almost every PISA item. The present analysis found that if one averages across the ACT and WorkKeys assessments, students have about 55 to 60 percent of the items with real-world context involved. The SAT II, on the other hand, is somewhat more guarded in departing from items that reflect only mathematical contexts. About 20 percent of the SAT II items involve context, compared to about 30 percent of ACT Assessment items. The balance provided for the PSAE by the ACT Assessment in conjunction with WorkKeys *Applied Mathematics* appears to give students an ample percentage of items with context. Hence, the PSAE is adequately assessing the goal of student ability to function in context-rich situations.

Table 6: Number and Percent of Items by Use of Real-World Context

	Items with Context	
	Number	Percent
ACT--Form 58B	18	30
ACT--Form 58E	19	32
WorkKeys--A07BB	33	100
WorkKeys--C01BB	33	100
SAT II-Level 1C	9	18
PISA--Math. Lit.	30	94

Computation

The next variable we considered was the proportion of items that required students to perform some aspect of computation in arriving at an answer. The computation might have been a mental calculation of a basic fact, an approximation, or the use of an algorithm that would have been difficult to complete without the aid of a hand calculator. This variable simply measured the presence or absence of such a requirement in the problems on each of the assessments studied. The results of the analysis of computation are shown in Table 7.

Table 7: Number and percentage of Items that Involve a Computation

	Items with Computation	
	Number	Percent
ACT--Form 58B	51	85
ACT--Form 58E	50	83
WorkKeys A07BB	33	100
WorkKeys C01BB	33	100
SAT II--Level 1C	40	80
PISA--Math. Lit.	19	59

A look at Table 7 shows that each of the tests, with the exception of the PISA assessment, requires students to perform some form of calculation in 80 percent or more of its items. The WorkKeys *Applied Mathematics* forms led the way, requiring a computation in every problem. The ACT Assessment forms required a computation in about 85 percent of their problems, and the SAT II examination called for some form of calculation in 80 percent of its items. The PISA assessment, drawing on more areas of content, only called for calculations in 59 percent of its problems. Clearly, in each case, with the possible exception of the PISA assessment, students are being called to use knowledge from the category of number sense and operations, whether or not that category is shown as being weighted heavily in the composition of main areas of content on the assessments. Parents of Illinois students do not need to worry that the basics of calculations are not being tested on the PSAE.

Calculator Usage

The data in Table 8 reflect whether a calculator might have been of some use in responding to the individual items on each of the assessments. The criterion applied in making this judgment for an individual item was whether or not the item required a calculation that went beyond the basic facts for the four operations of addition, subtraction, multiplication, and division with whole numbers. While the expectations that we hold for grade 11 students are higher than this, we established this level for making a judgment about whether a calculator might be of use to a student because we have seen this level of usage in classrooms and the basic-facts level was easy to enforce in rating the items on the various assessments.

Table 8: Number and Percent of Items where a Calculator Might be Used

	Calculator-Aided Items	
	Number	Percent
ACT--Form 58B	29	48
ACT--Form 58E	25	42
WorkKeys--A07BB	30	91
WorkKeys--C01BB	31	94
SAT II--Level 1C	20	40
PISA--Math. Lit.	6	19

The results show that the ACT Assessment forms and the SAT II were roughly equivalent in the potential effect that calculator use might have on students' responses, with the ACT Assessment being perhaps a bit more susceptible to impact from students' use of a calculator. Approximately 90 percent of the items on each WorkKeys *Applied Mathematics* assessment were open to influence by the use of calculators. On the PISA examination, on the other hand, only about 20 percent of the items were open to influence by calculator use. Again, this was partly because the PISA assessment was more balanced across the content areas and because it placed a heavier emphasis on conceptual items than on procedural items.

Multistep Thinking

If an assessment is to involve a student in significant problem solving, its items must require more than a simple one-step solution of its problems. A real-world problem—that is, a problem that reflects life—usually requires the blending of information and often the making of connections between disciplines to reach a solution.

Analysis of the composition of the assessments studied in this variable shows that the ACT and SAT II assessments were relatively equal in their employment of problems requiring two or more steps. About 82 to 87 percent of the items on these tests required more than one step to solve. The WorkKeys problems were a bit easier in terms of the demand defined by number of steps. Here only about 73 percent of the items required two or more steps. The PISA items were judged the easiest from this standpoint. Our analysis found only about half of the items, 53 percent, required more than one step.

Table 9: Number and Percent of Items Involving Single and Multistep Reasoning

	Single Step		Two or More Steps	
	Number	Percent	Number	Percent
ACT--Form 58B	8	13	52	87
ACT--Form 58E	9	15	51	85
WorkKeys--A07BB	9	27	24	73
WorkKeys--C01BB	9	27	24	73
SAT II--Level 1C	9	18	41	82
PISA--Math. Lit.	15	47	17	53

Combining the ACT and WorkKeys assessments leads to an overall level of about 82 percent of the items involving two or more steps for their solution. This is a respectable level of demand for students.

Representation

The statement or presentation of a problem can be placed in a graphical, tabular, symbolic, or verbal format. Each of these approaches, or some combination of them, potentially requires students to be able to translate the information into another format and potentially to use another representational form to either process the transformed information or to provide an answer to the problem posed.

Table 10: Number and Percentage of Items that Involve Interpreting a Representation

	Items with Representations	
	Number	Percent
ACT--Form 58B	22	37
ACT--Form 58E	17	28
WorkKeys--A07BB	7	21
WorkKeys--C01BB	6	18
SAT II--Level 1C	17	34
PISA--Math. Lit.	32	100

Table 10 presents the finding of the analysis of the use of representations in the presentation of items. Here there was a greater variation among the tests, even between different forms of the same assessment, in the use of representations. On average, the ACT Assessment forms employed some type of representation in about 33 percent of their items. The SAT II weighed in at 34 percent of its items using representations. PISA items had some type of representation in every item. The WorkKeys *Applied Mathematics* forms, on the other hand, with their high percentage of number and operation items, employed representations in only about 20 percent of their problems. It appears that the standard set by the ACT Assessment and SAT II examinations is appropriate. When a WorkKeys *Applied Mathematics* form and ACT Assessment form are combined to make up a given administration of the PSAE, the total percentage of items making use of a representation is about 55 percent of the items. Again, this appears to be a reasonable level of representations in the problems, especially given the timed nature of the test.

Table 11 provides a look at the various forms of representations employed in the tests we analyzed. An examination of the results suggests that there is some consistency within each of the individual tests in the representations used in items presented to students.

Table 11: Type of Representation in Items Having a Representation of Information*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
ACT--Form 58B	7	3	1	5	1	-	-	3	2	-
ACT--Form 58E	9	3	1	1	1	-	-	2	-	-
WorkKeys--A07BB	1	-	-	3	-	-	1	2	-	-
WorkKeys--C01BB	-	-	-	3	1	-	-	2	-	-
SAT II--Level 1C	14	-	-	1	1	-	-	-	1	-
PISA--Math. Lit.	8	1	-	-	12	-	-	3	3	5

*1-Geometric Figure or Drawing; 2-Algebraic/Functional Graph; 3-Number Line; 4-Data Table; 5-Statistical Graph; 6-Probability Situation; 7-Scale or Proportion Drawing; 8-Sketch Depicting Measurements of an Objects or Setting; 9-Depiction of an Algebraic Pattern; 10-Photograph

The ACT Assessment uses the widest variety of forms of representation. Each of the ACT Assessment forms that we reviewed used six or more different types of representation across its items. The WorkKeys *Applied Mathematics* forms used three or fewer types of representations. The SAT II used four different types, with most of them being clustered in geometric figures. The PISA assessment spread its items out over six different categories of representation. In the ACT and SAT II assessments, the most prevalent representation was a geometric figure or drawing. In the WorkKeys forms, the most prevalent representation was a data table. In the PISA assessment, the most prevalent representation was a statistical graph. The ACT and WorkKeys assessments together provide a wide range of representations for students to interpret. This range is acceptable for assessing students' problem-solving abilities.

Summary

This section presents a summary of our findings as well as some questions and issues that were raised during the analysis. First, in comparison with the SAT II and the PISA examinations, one of the components of the PSAE, WorkKeys *Applied Mathematics*, appears to have a heavy emphasis on the content area of number and operations, more than is necessary for students in grades 10 and 11. Although this is somewhat more balanced when the ACT Assessment Mathematics Test is included to form the PSAE, it raises the question of whether there are other ways to test students' number skills. In other words, can students' basic skills in number and operation be assessed through problems involving measurement, geometry, and algebra? If so, this may help to create a better balance across content areas.

A second major finding has to do with assessing the "Applications of Learning" as found in the Illinois Learning Standards (ISBE, 1997). These applications include solving problems, communicating, using technology, working in teams, and making connections. The components of the PSAE appear to do an adequate job of assessing problem-solving ability. This conclusion is based on the analysis of cognitive demand, multistep thinking, and representation, as reported in this analysis. It was found that the balance between routine and nonroutine problems was respectable on both the ACT Assessment and WorkKeys *Applied Mathematics*. In addition, there were a large number of items that required multiple steps or that required the interpretation of some representation. All of these aspects contribute to assessing problem-solving ability. The only aspect of problem solving that is not assessed by the PSAE is students' ability to support answers through reasoning and evidence. The PSAE assesses communicating, which is defined as expressing and interpreting information and ideas, only adequately. All test items require students to interpret the given information and identify the correct response. However, the multiple-choice format of the items does not provide students the opportunity to formulate their own responses and communicate their findings in writing. As noted previously, short-answer and extended-response items would provide such opportunities and would produce more valuable information on student communication skills in mathematics situations.

Based on analyses of problem context and representation, we concluded that the PSAE appears to address the area of making connections to a respectable degree. As indicated in our analysis, the WorkKeys items are all based on real-world applications. In addition, more than 30 percent of the ACT Assessment items contain context of some form. Both the ACT Assessment and WorkKeys also contain an appropriate number and variety of items with representations. These types of items help assess students' ability to make connections within mathematics and in settings beyond the classroom. As with problem solving, the addition of extended-response items will provide yet another opportunity for students to recognize and apply connections to the mathematics they have learned. The learning applications of using technology and working in teams were not appropriate for analysis.

In terms of cognitive demand, both components of the PSAE were found to be well in balance with the other examinations reviewed for this analysis. And finally, we judged that calculator use on computation items may be a bit higher than imagined, because of the widespread use of calculators for all levels of calculations at the high school level. In other words, the number of problems in which a calculator would likely be used is a bit high, but likely consistent with the students' high school experiences. It might be informative to take a closer look at what is actually being assessed by items for which a calculator is likely to be used. In other words, are the items actually assessing student understanding of mathematics concepts and procedures? Or, are these items testing only inappropriate, but accurate, use of the calculator?

Overall, the two components of the PSAE, taken together, assess a wide range of mathematical abilities. Of the two components, the ACT Assessment Mathematics Test appears to be a better constructed assessment in terms of its balance of content, computation, cognitive demand, and representation. The WorkKeys *Applied Mathematics* is less balanced in content (heavy in number and operation) and less balanced in variety of representations. *Applied Mathematics* certainly contains a greater number of items placed in real-world context than does the ACT Assessment, but this does not guarantee a thorough assessment of mathematics understanding.

Related to the recommendations listed in this summary, several issues and questions will be important to consider:

1. What role can more open-ended items play in assessment of Illinois students?
2. What is the role of the calculator on standardized tests such as the PSAE? How can either the testing procedures or the structure of the tests be altered to ensure an appropriate measure of both students' knowledge of mathematics and their ability to use technology in appropriate and powerful ways?
3. Do the context-rich items of WorkKeys *Applied Mathematics* provide enough of a good balance in terms of the other variables analyzed? If not, what other instruments are available to replace this or supplement the use of *Applied Mathematics* as part of the PSAE?

References

ACT, Inc. (2000). WorkKeys: Applied Mathematics-Helping to Build a Winning Workforce. Iowa City, IA: Author.

ACT, Inc. (2001). Contents of the Tests in the ACT Assessment. Iowa City, IA: Author.

Braswell, J. S., Lutkus, A. D., Grigg, W. S., Santapau, S. L., Tay-Lim, B. S., & Johnson, M. S. (2001). The Nation's Report Card: Mathematics 2000. Washington, DC: National Center for Education Statistics.

Burrill, G., Dossey, J., Paulson, D., & Webb, N. (1997). Setting Higher Sights: A Need for More Demanding Assessments for U.S. Eighth Graders. Washington, DC: American Federation of Teachers.

Dossey, J. A. (1996) Mathematics Examinations. In E. D. Britton & S. A. Raizen (eds.) Examining the Examinations: An International Comparison of Science and Mathematics Examinations for College-Bound Students (pp. 165-195). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Dossey, J. A., & Lindquist, M. M. (2001). External Review of the ISAT and other Standardized Mathematics Tests. Technical Report Prepared for the Assessment Division of the Illinois State Board of Education.

Dossey, J., Peak, L., & Nelson, D. (1997). Essential Skills in Mathematics: A Comparative Analysis of American and Japanese Assessments of Eight-Graders. Washington, DC: National Center for Education Statistics.

Gandal, M., & Dossey, J.A. (1997). What Students Abroad Are Expected to Know About Mathematics. Washington, DC: American Federation of Teachers.

Illinois State Board of Education. (1997). Illinois Learning Standards. Springfield, IL: Author.

Illinois State Board of Education. (1998). Item and Test Specifications. Springfield, IL: Author.

McLaughlin, D., Dossey, J., & Stancavage, F. (1997). Validation Studies of the Linkage Between NAEP and TIMSS Fourth and Eight Grade Mathematics Assessments. Washington, DC: Educational Statistical Services Institute.

National Assessment Governing Board. (1994). Mathematics Framework for the 1996 and 2000 National Assessment of Educational Progress. Washington, DC: Author.

National Assessment Governing Board. (2001). NAEP Mathematics 2005. Washington, DC: Author.

National Council of Teachers of Mathematics. (2000). Principles and Standards for School Mathematics. Reston, VA: Author.

Nohara, D., & Goldstein, A. (2001). A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA). NCES Document 2001-07. Washington, DC: National Center for Education Statistics.

Organization for Economic Cooperation and Development. (2000). Measuring Student Knowledge and Skills: The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy. Paris, France: OECD.

The College Board. (1998). Real SATII Subject Tests: Math IC/Math IIC. New York, NY: Author.

The College Board. (2001). Taking the SAT II Subject Tests: 2001-2002. New York, NY: Author.

U.S. Office of Education. (2001). Outcomes of Learning. Washington, DC: National Center of Education Statistics.

Addendum to the External Review of the PSAE Mathematics Test

To: ISBE Student Assessment Division
From: John A. Dossey and Sharon S. McCrone
Re: Addendum to External Review of the Prairie State Achievement Examination Mathematics Test (Dossey & McCrone, 2002)
Date: November 20, 2002

Pursuant to your request that we revisit our analysis of the *ACT Assessment* and *WorkKeys Assessment* relative to the fit of these instruments to the Illinois Learning Standards (1999), we submit the following report.

Summary

The analysis of two Prairie State Achievement Examination (PSAE) forms and additional released items—the forms contained in the previous analysis and the released form and items added in this study—indicates that the PSAE compares well with other major assessments. In fact, the PSAE provides a balanced assessment that comes closer to adequately assessing the Illinois Learning Standards than does either The College Board's SAT II, Level IC examination, an achievement test aimed at students who should have completed three years of high school mathematics or the PISA mathematics literacy assessment (Dossey & McCrone, 2002). The present analysis, see Table 3, indicates that the merged content-area means of the PSAE (merged data from the ACT Assessment and WorkKeys *Applied Mathematics*) fall within the ranges for similar content-area means of state assessments from across the United States with the sole exception of Data/Chance. With a minor change in the balance of items in the areas of Number and Operation and Data/Chance, the balance could easily be made to fall totally within the ranges. The observed percentages are also quite reasonable relative to the National Assessment of Educational Progress (NAEP) 2005 percentage targets as we discuss later in this addendum (National Assessment Governing Board (NAGB), 2001).

The balanced content of the PSAE, coupled with its excellent balance of cognitive demand across the items, gives the PSAE a range of items that adequately assess *all students*. In like manner, the PSAE has a solid balance of context and noncontext items and of computation/calculator active items. The PSAE also has a solid balance of items that require conceptual and procedural knowledge in mathematics. Finally, the PSAE has a quite acceptable percentage of items requiring students to make an interpretation of a representation as part of their response. The data in Tables 9 and 10 reflect that about 25 to 30 percent of the items make use of some representation. This indicates that the PSAE requires students to make use of a variety of ways of representing information in addition to verbal and symbolic representations. This use of varied representations is in line with the emphasis on representation in the National Council of Teachers of Mathematics (NCTM) recommendations for the secondary mathematics curriculum.

As such, the PSAE is a broad and demanding assessment of secondary school mathematics. Its breadth is comparable to that found in other state assessments and is in line with the assessment guidelines of both the Illinois State Board of Eduction (ISBE) and NAGB with the exception of Data/ Chance, a difference that

can be easily remedied with a little more emphasis on Data/Chance and a slight decrease in the Number and Operation items.

The Process

At the request of ISBE, we reexamined our analyses of this past summer and expanded the analysis to include data from the released version of the ACT Assessment (Form 57B) and the 15 example items from WorkKeys *Applied Mathematics* contained in *Prairie State Achievement Examination: Teachers Handbook 2001-2002* (ISBE, 2001). Thus, our reanalysis is based on the items contained in the following forms of the assessments that make up the PSAE mathematics test:

- Mathematics Test, *ACT Assessment*, Form 58B, ACT, Inc., 1999.
- Mathematics Test, *ACT Assessment*, Form 58E, ACT, Inc., 1999.
- Mathematics Test, *ACT Assessment*, Form 57B, ACT, Inc., n.d.
- Applied Mathematics Test, *WorkKeys Assessment*, Form A07BB, ACT, Inc., 2001.
- Applied Mathematics Test, *WorkKeys Assessment*, Form C01BB, ACT, Inc., 2001.
- Applied Mathematics Test, *WorkKeys Assessment*, Example Items, ACT, Inc., n.d.

We used the National Assessment of Educational Progress (NAEP) framework and the Illinois Learning Standards as guides for our reexamination of the data (NAGB, 2001). The NAEP 2005 goals, for instance, suggest a specific balance of items for student assessment as can be seen in the middle column of Table 1. The Illinois Learning Standards, on the other hand, do not suggest a specific balance of items on which to assess students. Thus, we have used the NAEP framework and other sources to help determine a suitable balance of assessment items. It should also be noted that the five content areas of NAEP and the Illinois Learning Standards are very representative of the mathematics content areas found in the National Council of Teachers of Mathematics' *Principles and Standards for School Mathematics* (NCTM, 2000) and the learning standards of almost all of the other states (Dossey, 2002). Data from the Dossey 2002 study indicated that states varied somewhat in the balances they gave to the five learning areas.

Table 1: Recommended percentages and assessment emphases on grade 12 mathematics assessments

Content Area	NAEP 2005 Recommendations	Ranges of State Emphases
Number sense	10%	14–40
Measurement	30%[a]	11–25
Geometry		9–25
Algebra	35%	8–35[b]
Data/Chance	25%	14–34

[a] The recommendation is that in 2005 the total combined geometry and measurement items make up 30 percent of the questions. [b] The state of California's high school test is an outlier in the set of state examination in that it is made up of 100 percent algebra items.

Analysis of the various forms of the PSAE components with respect to these five areas is shown in Table 2. In addition to breaking down the assessment forms into the five major areas of the Illinois Learning Standards (1997), two of the areas, Algebra and Geometry, are broken down into finer components. This finer breakdown ensures that the assessments have some balance between conceptual and applied/procedural aspects in these two major areas of the secondary mathematics curriculum. Note also that the number of items in a category is sometimes given in decimals. This occurs where an item spans one or more categories, and it was impossible to place the item in a specific category. In these cases, the count was equally prorated across the possible categories.

Table 2: Number and percentage of items relative to the Illinois Learning Standards.

	ACT Form 58B		ACT Form 58E		ACT From 57B		WorkKeys A07BB		WorkKeys C10BB		WorkKeys Example	
	#	%	#	%	#	%	#	%	#	%	#	%
NUMBER	8	13	10	17	9.16	15	22	67	20	61	7.5	50
MEASUREMENT	8	13	7	12	6.83	11	9	27	10	30	6.5	43
GEOMETRY	(11)	(19)	(14)	(21)	(14.8)	(25)	(0)	(0)	(0)	(0)	(0)	(0)
Concepts	7	12	3	5	4.33	7	0	0	0	0	0	0
Relations	4	7	11	16	10.50	18	0	0	0	0	0	0
ALGEBRA	(27)	(45)	(24)	(40)	(27)	(45)	(0)	(0)	(0)	(0)	(0)	(0)
Patterns & Variables	13	22	12	20	15	25	0	0	0	0	0	0
Relations/ Representation	14	23	12	20	12	20	0	0	0	0	0	0
DATA/CHANCE	(6)	(10)	(5)	(8)	(2)	(3)	(2)	(6)	(3)	(9)	(1)	(7)
Data Analysis	4	7	3	5	1	2	2	6	3	9	1	7
Probability	2	3	2	3	1	2	0	0	0	0	0	0

Table 3 shows the percentage of items in each of the five major learning areas for the PSAE components reviewed. In addition, the table allows for comparison of each form against the NAEP 2005 ranges and comparison of a combined average of all PSAE forms against the NAEP ranges (NAGB, 2001). This final comparison shows the balanced average percentage of the five content areas found by merging the various ACT and WorkKeys forms as a model for the PSAE. Comparing this to the NAEP and survey ranges from Table 1, we found that the PSAE averages fall within all of the state ranges except for items from Data/Chance. In this content area, the PSAE average percentage is beneath the lower bound of the range interval. In comparison to the NAEP ranges, the ACT Assessment average matches up well with the exception of the Data/Chance area. The WorkKeys forms fall above the range interval in Number and Measurement and beneath it in Geometry, Algebra, and Data/Chance.

Table 3: Percent of PSAE assessment areas by NAEP and state ranges

	58B	58E	57B	**ACT Average**	NAEP	A07BB	C10BB	Example	**WorkKeys Average**	NAEP	Merged Mean	Within Range
Number	13	17	15	**15**	10	67	61	50	**60**	10	31	YES
Measurement	13	12	11	**12**	30	27	30	43	**33**	30	19	YES
Geometry Concepts Relations	19	21	25	**22**		0	0	0	**0**		14	YES
Algebra	45	40	45	**43**	35	0	0	0	**0**	35	28	YES
Data/chance	10	8	3	**7**	25	6	9	7	**7**	25	7	NO

Based on these comparisons, the PSAE does a credible job of matching up to the NAEP and state ranges. The addition of a few more Data/Chance items and the deletion of several Number and Operation items would bring the PSAE closer to the NAEP balance.

In addition to item analysis by content areas, we compared ACT Assessment Mathematics (Form 57B) and the WorkKeys sample items from the *Teacher's Handbook* to other forms of these same assessments along other pertinent variables. These include: cognitive demand, use of real-world context, amount of computation, possibility of calculator use by students, multistep reasoning, and use of representations.

The expanded analysis of the ACT Assessment and WorkKeys forms indicated that our cognitive demand comparisons did not change significantly from the original report (Dossey & McCrone, 2002). That is, the PSAE seems to have a nice range of items at each of the levels of cognitive demand. This information is shown in Table 4.

Table 4: Number and percentage of items by cognitive demand categories

		Number of Items		Percentage of Items	
		Routine	Nonroutine	Routine	Nonroutine
ACT Form 58B	Simple	23	16	38	27
	Complex	16	5	27	8
ACT Form 58E	Simple	19	20	32	33
	Complex	11	10	18	17
ACT Form 57B	Simple	8	19	13	32
	Complex	24	9	40	15
WorkKeys A07BB	Simple	14	7	42	21
	Complex	7	5	21	15
WorkKeys C01BB	Simple	14	10	42	30
	Complex	4	5	12	15
WorkKeys Examples	Simple	4	3	27	20
	Complex	4	4	27	27

The analysis of the two new forms with respect to the use of real-world contexts is shown in Table 5. The percentages are essentially the same as for the forms analyzed earlier. This percentage is quite acceptable given the time-bounded assessment format.

Table 5: Number and percentage of items by use of real-world context

	Items with Context	
	Number	Percentage
ACT—Form 58B	18	30
ACT—Form 58E	19	32
ACT—Form 57B	18	30
WorkKeys—A07BB	33	100
WorkKeys—C01BB	33	100
WorkKeys—Examples	15	100

Computation is a major facet of applied mathematical problem solving. Table 6 shows the percentage of items requiring examinees to perform a computation of any type in the completion of the item. This comparison shows a slight decrease in the percentage of items on Form 57B that call for a calculation.

Table 6: Number and percentage of items that involve a computation

	Items with Computation	
	Number	Percentage
ACT—Form 58B	51	85
ACT—Form 58E	50	83
ACT—Form 57B	42	70
WorkKeys A07BB	33	100
WorkKeys C01BB	33	100
WorkKeys Examples	15	100

The results of an analysis of items for which student performance might be assisted with the use of a calculator are reported in Table 7. This analysis showed a slight decrease in the percentage of items on Form 57B where a calculator might be of some assistance for students. This parallels the slight decrease in the number of calculation items shown in Table 6. This decrease is probably not a concern in an overall analysis of the test, given the large number of Number and Operation items found in the WorkKeys assessment.

Table 7: Number and Percentage of Items for which a Calculator Might be Used

	Calculator-Aided Items	
	Number	Percentage
ACT—Form 58B	29	48
ACT—Form 58E	25	42
ACT—Form 57B	21	35
WorkKeys—A07BB	30	91
WorkKeys—C01BB	31	94
WorkKeys—Examples	12	80

The decrease in the number of calculation items noted in Table 6 also carries over into the analysis of multistep reasoning items as reflected in Table 8.

Table 8: Number and percentage of items involving single and multistep reasoning

	Single Step		Two or More Steps	
	Number	Percentage	Number	Percentage
ACT—Form 58B	8	13	52	87
ACT—Form 58E	9	15	51	85
ACT—Form 57B	19	32	41	68
WorkKeys—A07BB	9	27	24	73
WorkKeys—C01BB	9	27	24	73
WorkKeys—Examples	4	27	11	73

Table 9 contains the data showing the number and percentage of items containing a representation that provides further information to the student. These representations were noted only when they were different from the usual printed instructions or equations. Such representations could consist of a geometric figure or drawing, an algebraic/functional graph, a number line, a data table, a statistical graph, a probability situation, a scale or proportion drawing, a sketch depicting measurements of objects or setting, a depiction of an algebraic pattern, or a photograph. The data in Table 9 show a great deal of consistency when the new forms are added to the forms previously analyzed. Table 10 contains the data showing the types of representations that were found in the forms analyzed.

Table 9: Number and percentage of items that involve interpreting a representation

	Items with Representations	
	Number	Percentage
ACT—Form 58B	22	37
ACT—Form 58E	17	28
ACT—Form 57B	19	32
WorkKeys—A07BB	7	21
WorkKeys—C01BB	6	18
WorkKeys—Examples	4	27

Table 10: Type of representation in items having a representation of information*

	1	2	3	4	5	6	7	8	9	10
ACT—Form 58B	7	3	1	5	1	-	-	3	2	-
ACT—Form 58E	9	3	1	1	1	-	-	2	-	-
ACT—Form 57B	14	3	1	1	-	-	-	-	-	-
WorkKeys—A07BB	1	-	-	3	-	-	1	2	-	-
WorkKeys—C01BB	-	-	-	3	1	-	-	2	-	-
WorkKeys—Examples	1	-	1	1	1	-	-	-	-	-

*1-Geometric Figure or Drawing; 2-Algebraic/Functional Graph; 3-Number Line; 4-Data Table; 5-Statistical Graph; 6-Probability Situation; 7-Scale or Proportion Drawing; 8-Sketch Depicting Measurements of an Objects or Setting; 9-Depiction of an Algebraic Pattern; 10-Photograph

The data in these foregoing tables reflect our analysis of the additional forms provided by ISBE. Combining this information with that developed in the analysis provided last summer indicates that the PSAE provides a solid assessment that falls within both the Illinois Learning Standards and the NAGB content guidelines (ICTM, 1997; NAGB, 2001) and that adequately assesses the Illinois Learning Standards.

References

ACT, Inc. *ACT Assessments, Forms 57B and 57E*. Personal communications with ACT, Inc., as part of the original analysis in summer, 2002.

ACT, Inc. *WorkKeys Assessments Forms A07BB and C01BB.* Personal communications with ACT, Inc., as part of the original analysis in summer, 2002.

Dossey, J.A. (2002). *Survey of State Learning Standards and Assessment Formats.* Unpublished study carried out for Educational Testing Service, Summer, 2002.

Dossey, J.A., & McCrone, S. S. (2002). External Review of the Prairie State Achievement Examination Mathematics Test. In ACT/Illinois State Board of Education. *Prairie State Achievement Examination, Technical Manual*, pp. 327–344. Springfield, IL: Illinois State Board of Education.

Illinois State Board of Education. (1997). *Illinois Learning Standards.* Springfield, IL: Illinois State Board of Education.

Illinois State Board of Education. (2001). *Prairie State Achievement Examination: Teacher's Handbook: 2001-2002.* Springfield, IL: Illinois State Board of Education.

National Assessment Governing Board. (2001). *National Assessment of Educational Progress Mathematics Framework: 2005*. Washington, DC: NAGB.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM.