Prairie State Achievement Examination

Technical Manual

2006 Testing Cycle

ACT and the Illinois State Board of Education

Table of Contents

List of Figures

List of Tables

Preface

This manual documents the technical characteristics of the 2006 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 Test.® 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 reliability, standard errors of measurement, 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 2006 spring administration of the PSAE and Chapter 9 provides the results of the grade 12 retake in the fall of 2006.

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 only 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.

The PSAE is administered statewide during a two-day period each year in April. Students took the PSAE for the first time in April 2001. In alignment with the Illinois Learning Standards and in accordance with state law (105 ILCS 5/2-3.64), the 2006 PSAE assesses three academic subjects: reading, mathematics, and science.

Components of the PSAE

The PSAE comprises assessments from three sources: (1) the ACT® Test, which includes tests in English, mathematics, reading, and science; (2) an ISBE-developed science assessment; 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 Tests		Components of Each PSAE Test
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 two purposes: (1) to measure students' progress toward meeting the Illinois Learning Standards for state and federal accountability requirements and (2) to recognize the achievement of individual students who earn a Prairie State Achievement Award for excellent performance.

Population Served by the PSAE

All eligible grade 11 public-school students take the PSAE. 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 state's alternate assessment. Students with limited English proficiency (LEP) must take the PSAE, except for students who have been in state-approved bilingual education programs for less than three cumulative academic years and whose lack of English as determined by an English language proficiency test would keep them from understanding the PSAE test questions.

In April 2006, the PSAE was administered in Illinois in grade 11. Table 1.2 presents the demographic characteristics of the grade 11 students tested in 2006.

Table 1.2: Demographic characteristics of grade 11 students taking the spring 2006 PSAE (Reported as percentages)

Gender	**Percent**
Female	51
Male	49
No response	0
Race/Ethnicity	
African American/Black	16
Caucasian American/White	64
American Indian/Alaskan Native	0
Hispanic (non-black)	12
Asian American/Pacific Islander	4
Multiracial/Ethnic	1
No response	4

Administration of the PSAE

The PSAE is administered annually over a two-day period in April. Day 1 consists of the ACT Test and Day 2 consists of the two WorkKeys assessments and the ISBE-developed science test. Table 1.3 presents the April 2006 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.

An opportunity to retake the full PSAE is offered in October (following spring testing) for grade 12 students who wish to improve their PSAE scores or take the PSAE for the first time if they missed the spring administrations. Day 1 of the PSAE retake (the ACT Test) is the October national ACT test date. The state provides vouchers for schools to give to students to pay their registration for the October ACT. Day 2 of the PSAE retake is scheduled for Tuesday in the week following the Saturday Day 1 administration. Day 2 is administered during an in-school session. It is critically important that the PSAE be administered under secure, standardized conditions. To ensure that the educators named as test supervisors or back-up test supervisors for their schools were trained in conducting a standardized administration, ISBE and ACT conducted 16 training workshops throughout the state in February and March 2006. ISBE required that both the test supervisor and back-up test supervisor from every school that would serve as a PSAE test site, including those individuals who had previously served as ACT national test supervisors, attend one of these workshops before secure test materials could be shipped to schools. All schools that were administering the PSAE complied with the requirement.

At the workshops, participants were provided with a packet of materials to assist them with planning and carrying out the test administration. The information that was provided included planning for the test days, selecting and training staff members to serve as room supervisors and proctors, maintaining the security of test materials at all times, administering the tests under standardized conditions, ensuring exact timing of the tests, handling testing irregularities, and providing accurate written documentation of test day procedures. Workshop leaders explained the standardized testing requirements for the PSAE in detail.

The workshop packet included copies of answer documents and other test-related forms so that testing

Table 1.3: PSAE 2006 test-administration schedule

Day 1	ACT English Test – 45 minutes (75 questions) ACT Mathematics Test – 60 minutes (60 questions) *[15-minute break]* ACT Reading Test – 35 minutes (40 questions) ACT Science Test – 35 minutes (40 questions)
Day 2	ISBE-Developed Science – 40 minutes (40 questions) *[15-minute break]* WorkKeys *Applied Mathematics* – 45 minutes (33 questions) *[5-minute break]* WorkKeys *Reading for Information* – 45 minutes (33 questions) *[15- to 60-minute break as required by school schedules]*

staff could become familiar with them before the test days. In addition, each workshop participant received a copy of the ACT *Supervisor's Manual for State Testing* for Day 1 testing and the *Day 2 Prairie State Achievement Examination Supervisor's Manual of Instructions* for Day 2 testing. 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 6:30 a.m.

Accommodations for Students with Disabilities

Special testing services (accommodations for students with disabilities) are available for both Day 1 and Day 2 of the PSAE. ACT must approve accommodations for Day 1; school staff must submit a request form for each eligible student who wishes to request an accommodation based on determinations documented in their IEP or Section 504 Plan. ACT staff review the information provided and inform the school of approved accommodations for the ACT. School staff may use the ACT approval as a guideline for accommodations to be provided for Day 2, but the decision regarding Day 2 accommodations is ultimately a local one based on accommodation determinations documented in a student's IEP or Section 504 Plan.

Examples of test accommodations include the following:

1. Extended time
2. Large-type/print test booklet
3. Testing over more than one day
4. Stop-the-clock breaks with standard time
5. Alternate test formats, including Braille, audiocassettes, and reader scripts

To be considered for test accommodations for the ACT, students with disabilities must meet ALL of the following requirements:

- Have a 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.
- Have a current IEP or 504 Plan that specifies extended time.
- Be allowed extended time or other test accommodations for regular school work. For school tests, the student must currently receive the accommodations being requested because of the disability, including extended time.

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

Accommodations for LEP students are not available on the PSAE. LEP students who are enrolled in state-approved bilingual education programs for less than three cumulative academic years and whose lack of understanding of English, as determined by an English language proficiency test, would keep them from understanding PSAE test questions are required to take the Illinois Measure of Annual Growth in English (IMAGE) test in reading and mathematics as an alternative state assessment.

Chapter 2
Evidence of Validity of 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 its test-development contractor, Harcourt.
- WorkKeys® tests in reading and mathematics, and
- The ACT® Test.

The Prairie State Achievement Examination 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 on the following page 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 Test, two WorkKeys assessments, and the ISBE-developed test covered the necessary content, ISBE conducted reviews that compared the contents of these tests with the Illinois Learning Standards.

ISBE first reviewed the ACT Test and a study that ACT had previously done that compared the ACT Test to the Illinois Learning Standards. ISBE then 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 Test, 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.

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), including literature representative of various societies, eras, and ideas (2A, B). Ability to evaluate and use information from various sources to answer questions, solve problems, and communicate ideas (5A–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

The ACT Test's Match to the Illinois Learning Standards

The ACT Test 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 2007 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 define what teachers should be teaching, the ACT has a direct and close relationship to state standards.

ACT staff also did a match between the ACT Test and the standards of more than 30 states, including the Illinois Learning Standards. ISBE reviewed ACT's study comparing the skills assessed on the ACT with the Standards. The 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 that which is 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 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* (February 1999) and *Comparison of the Illinois Learning Standards to the ACT Assessment Standards for Transition* (February 2000).

To conduct its own review of the relationship of the Illinois Learning Standards to the ACT Test, ISBE convened meetings of Illinois educators who were engaged in instruction aligned with the Illinois Learning Standards to review the match between the ACT Test and the Illinois Learning Standards. The results of this review also showed that there is substantial agreement between the ACT Test 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 and everyday life 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 Springfield on February 26, 2000, with 25 high school mathematics teachers. The science meeting was held in Champaign 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). They concluded that, in conjunction with the ISBE-developed writing assessment, the ACT English Test matches the Illinois Learning Standards under 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 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 workplace 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. Appendix B contains the detailed results of these reviews.

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 2006, 2006a, and 2006b).

Based on all reviews of the alignment between the PSAE components and the Illinois Learning Standards, ISBE concluded that although a few weaknesses are noted, overall the PSAE does a fairly good job at covering the Illinois Learning Standards in Reading, Mathematics, and Science.

Additional Evidence of the Validity of the 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 determine whether the ACT Test and WorkKeys assessments could be used appropriately 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. The evaluation process was undertaken to ensure that the program would work to the advantage—not disadvantage—of students.

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 three subjects: reading, mathematics, and science. 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 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. Workshops like these continue to be conducted each year Those who attend these workshops are expected to train teachers and administrators who were unable to attend. The teachers and administrators, in turn, are expected to share this information with students and their parents. Workshop participants are strongly encouraged to familiarize all students with the PSAE (for example, test dates and times, subjects assessed, types of tests, types of questions) and provide them opportunities to practice relevant test-taking skills (such as pacing and skipping questions that are too difficult). 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*. High schools also received information pertaining to the PSAE as a whole and the ISBE-developed assessments in November 2000, the *2000–2001 PSAE Overview and Preparation Guide for Writing, Science, and Social Science.* All of these materials were intended to familiarize teachers and students 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.

Finally, after scores were reported in August, workshops were held in September and November to review and evaluate the PSAE scores and to help high school teachers and administrators understand how to interpret and use their results. Students and their parents also received guides to score interpretation, *A Guide to Your Scores* (designed to accompany students' individual PSAE score reports) and *Using Your ACT Test Results.*

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 test supervisors or back-up test supervisors 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 four of the five subject areas that it measures. These scores have meaning and value within the statewide assessment context and should inform both instruction and individual improvement within the classroom setting. 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, and science. The ACT student report contains scores for each of the four ACT tests, seven 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.

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 selection, bias review, and test assembly. 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.

Precision. Point-biserial (i.e., item-test) correlations evaluate the extent to which an item distinguishes between less proficient and more proficient students. Reviewers usually omit items with a point-biserial of less than 0.30 and select items with the highest point-biserial.

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.

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.

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

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 that provide a rough balance across the middle ranks are selected, and items at the level of vocabulary or rote memory are usually omitted. Items are also examined to determine whether there is a reasonable distribution within tests of items across the standards: earth science, physical science, and life science.

Reliability of the ISBE-Developed Science Test

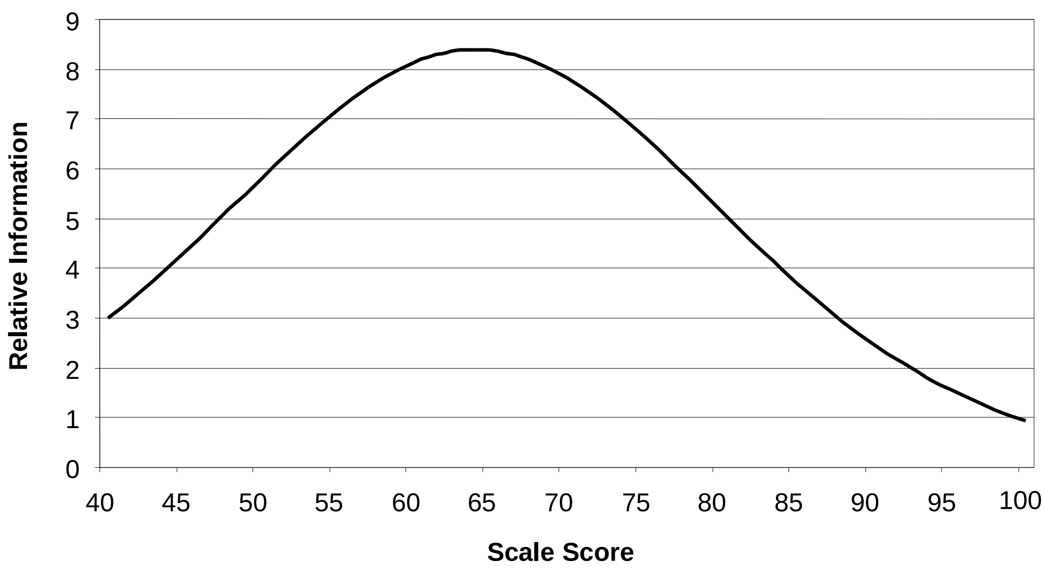
The reliability of a test reflects the degree to which scores are free from random errors of measurement. Test reliability indicates the extent to which differences in test scores reflect real differences in the ability being measured and, thus, the consistency of test scores across some change of condition, such as a change of test items or a change of time. Different reliability coefficients result from different changes in testing conditions. For example, test-retest reliability measures the extent to which scores remain constant over time. A low test-retest reliability coefficient means that an individual's scores are likely to shift unpredictably from one time to another.

Because the items used in achievement tests represent a sample from a much larger domain of items, the consistency of test scores across items is of particular interest. That is, How comparably will tests rank students if different sets of items from the same domain are used? Unless the rankings are very similar, it is difficult or impossible to make educationally sound decisions on the basis of test scores. This characteristic of test scores is most commonly referred to as *internal consistency*. For the ISBE-developed science test administered as part of the 2001 PSAE, the science score internal consistency value (coefficient alpha) reliability was 0.89 based on a sample size of 115,518.

The value is derived from the total test population. For well-constructed achievement tests, internal consistency reliability coefficients typically exceed 0.90, which is the case for the test reported here. 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 provide 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 PSAE science test.

Figure 2.1: 2001 PSAE science test information function



The amount of information at any point is directly related to the precision of the test. That is, precision is highest where information is highest. Conversely, where information is lowest, precision is lowest and ability is most poorly estimated. As with the ISATs, which also measure the Illinois Learning Standards, the information functions for ISBE-developed science test is highest near the points on the scales where the Meets Standards cut scores are located.

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 2001 PSAE ISBE-developed science test is 2.18. 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.

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 2.2 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.

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 2.3.

Table 2.2: Results of the 2001 calibration process – Science

Item	Difficulty	Sed	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

Table 2.3: PSAE scaling constants

	Slope	Intercept
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 first eight WorkKeys skill areas: *Applied Technology, Applied Mathematics, Listening, Locating Information, Observation, Reading for Information, Teamwork,* and *Writing.* Additional skill areas are in development.

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, there were no scales that could 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 (ACT, 1995) and receives written evaluations back from them.

Table 2.4 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.

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.

Table 2.4: Number of reviewers by type of review for the operational WorkKeys assessments

Assessment Title	Number of Content Reviewers	Number of Fairness Reviewers
Applied Mathematics	9	8
Reading for Information	13	8

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 7 requiring the most complex and Level 3 requiring the least complex mathematical concepts and calculations.

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.

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 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 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 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.

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 7 being the most complex and Level 3 the least complex.

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.

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 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 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 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.

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.

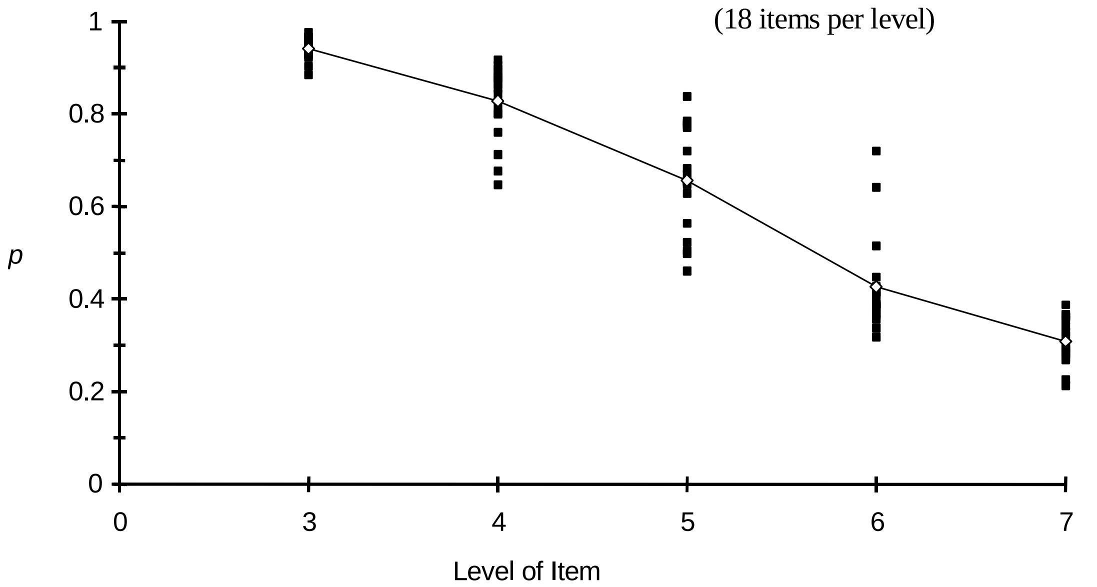
Alternate forms for the reading and mathematics skills, as well as for other WorkKeys multiple-choice tests, were administered to randomly equivalent groups of high school juniors and seniors in one state by spiraling forms within classrooms. This data collection process and the analyses that defined 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 2.5 on the following page. 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 2.5: 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

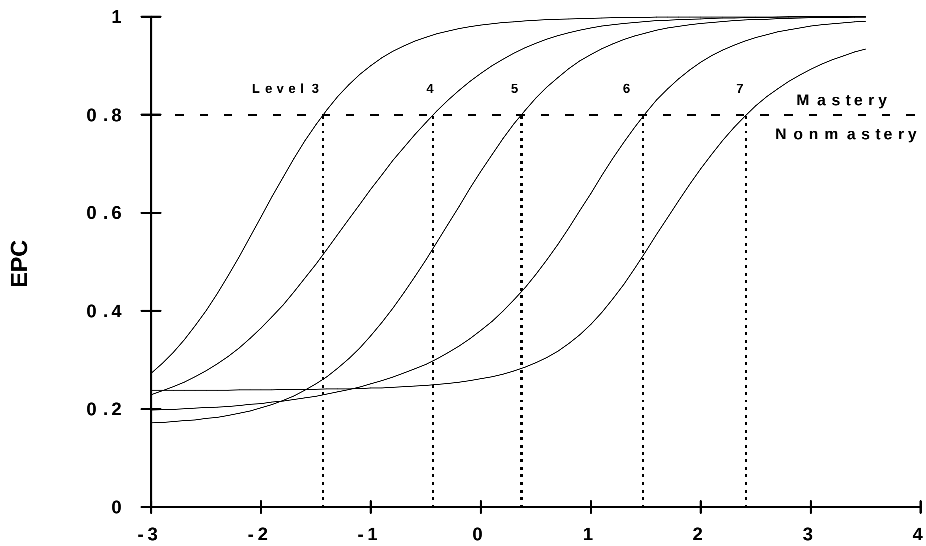
Figure 2.2: Item *p*-values (*p*) and mean item *p*-values (connected) by level of item on WorkKeys *Applied Mathematics* tests



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 2.3: ***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 2.6 shows the boundary thetas that define levels of WorkKeys skills. The lower boundary of Level 3 on the θ theta 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 2.6 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 2.7 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 2.8 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.

The θ corresponding to a cutoff score is referred to as a "form-specific cutoff θ." In Table 2.8, 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.

Table 2.6: 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 2.7: 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 2.8: 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

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 2.7) and their associated θs (Table 2.8) 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 2.9 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.

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.

Table 2.9: Summary statistics of level scores by form of *Applied Mathematics*

Level	Percentage		
	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

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 2.5 with the reliability of level scores reported in Table 2.10 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 35) 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 2.10 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.

Table 2.10: 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.

All of the percentages in the upper panel of Table 2.10 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 2.10. However, the predicted percentages were very close to the actual percentages shown in Table 2.10. 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).

The bottom panel of Table 2.10 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 2.11 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 2.11 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).

Table 2.11: 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.

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 2.12 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 2.12, 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 2.12 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.

Table 2.12: 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.

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).

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 2.6), is near the zero-mean of the *Applied Mathematics* θ distribution used to compute classification consistency and classification error. (The θ distribution for each skill is assumed to be standard normal as noted in Chapter 4.) 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 2.6), 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 2.11 and 2.12 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 Technical Handbook* (ACT, 2001) 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 chapter provides detailed information about the score distribution characteristics of a sample of examinees who took WorkKeys assessments in fall 1999 and spring 2000.

Description of the Sample

The sample contained 179,967 usable examinee records. The examinees included both high school students and adults (such as postsecondary students, employees, and job applicants). Although 48 states were represented in the sample, the representation was not equal as a few states administered more WorkKeys assessments than did many others.

The WorkKeys answer document contains a number of demographic questions used both in reporting scores and in research. Table 2.13 provides the n-count and percentages for selected demographic variables for this sample. Because some examinees did not give their personal information, the n-count may vary from variable to variable.

Score Distributions

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 or above each skill level are used to report the score distributions of the sample in this section.

Table 2.13: Descriptive statistics for selected 1999–2000 demographic variables

Demographic Variable	Number	Percentage
Gender		
Female	71,433	39.7
Male	81,826	45.5
No response	26,708	14.8
Race/Ethnicity		
African American/Black	22,158	12.3
American Indian/Alaskan Native	1,371	0.8
Asian American/Pacific Islander	1,748	1.0
Caucasian/White	92,811	51.6
Hispanic[a]	5,008	2.8
Other and Prefer Not to Respond, No Response	56,871	31.6
Program of Study		
College Preparation	31,365	17.4
General Education	25,845	14.4
Vocational Technical	13,964	7.8
Tech Prep	4,158	2.3
No Response	104,635	58.1
Highest Education Level (in grade level)		
7 and 8	1,647	0.9
9	19,508	10.8
10	9,820	5.5
11	11,994	6.7
12	39,399	21.9
High School Grad/GED	3,191	1.8
Postsecondary	3,206	1.8
No response	91,202	50.7
English is the language in which examinee communicates best		
Yes	107,778	59.9
No	2,044	1.1
No response	70,145	39.0

[a]Mexican American/Chicano; Puerto Rican; Cuban; Other Hispanic/Latino

Tables 2.12 and 2.13 contains the numbers and percentages of the examinees at or above each level of each operational WorkKeys assessment. The percentage at or above a level is an inverse cumulative frequency that can be used to compare groups within a skill area. The percentage at a particular level can be obtained by finding the difference in the percentages between this level and the level to the right (i.e., the next higher level). For example, in Table 2.14, for the total group, 68% of the examinees were at or above Level 4, and 42% were at or above Level 5. The difference was 26%, meaning that 26% of the examinees were at Level 4. For each skill area in Tables 2.14 and 2.15, the score distributions are compiled by total, gender, and ethnicity. It is important to note that 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 2.14: Numbers and percentages at or above each level of *Applied Mathematics* (Based on 1999–2000 data)

	Below Level 3	Level 3	Level 4	Level 5	Level 6	Level 7	Total Examinees
Total Group							
Number	15,018	26,777	34,791	36,216	14,768	3,868	131,438
Percent at or above	100	88.6	68.2	41.7	14.2	2.9	
Gender							
Female							
Number	5,547	11,352	15,260	15,254	4,946	949	53,308
Percent at or above	100	89.6	68.3	39.7	11.1	1.8	
Male							
Number	7,480	11,864	15,229	16,628	7,854	2,354	61,409
Percent at or above	100	87.8	68.5	43.7	16.6	3.8	
No Response							
Number	1,991	3,561	4,302	4,334	1,968	565	16,721
Percent at or above	100	88.1	66.8	41.1	15.1	3.4	
Race/Ethnicity							
African American/ Black							
Number	4,379	6,301	4,656	2,514	416	47	18,313
Percent at or above	100	76.1	41.7	16.3	2.5	0.3	
American Indian/ Alaskan Native							
Number	219	263	263	201	49	11	1006
Percent at or above	100	78.2	52.1	25.9	6.0	1.1	
Asian American/ Pacific Islander							
Number	142	248	322	362	175	66	1,315
Percent at or above	100	89.2	70.3	45.9	18.3	5.0	
Caucasian/White							
Number	4,738	10,376	17,507	21,187	9,575	2,612	65,995
Percent at or above	100	92.8	77.1	50.6	18.5	4.0	
Hispanic							
Number	535	917	1,080	799	221	44	3,596
Percent at or above	100	85.1	59.6	29.6	7.4	1.2	
Other, Prefer Not to Respond, No Response							
Number	5,006	8,672	10,963	11,153	4,332	1,089	41,215
Percent at or above	100	87.9	66.8	40.2	13.2	2.6	

Table 2.15: Numbers and percentages at or above each level of *Reading for Information* (Based on 1999–2000 data)

	Below Level 3	Level 3	Level 4	Level 5	Level 6	Level 7	Total Examinees
Total Group							
Number	16,015	15,173	50,211	35,865	21,448	3,667	142,379
Percent at or above	100	88.8	78.1	42.8	17.6	2.6	
Gender							
Female							
Number	4,491	5,650	22,713	16,767	9,673	1,497	60,791
Percent at or above	100	92.6	83.3	46.0	18.4	2.5	
Male							
Number	9,428	7,524	21,029	14,379	8,705	1,589	62,654
Percent at or above	100	85.0	72.9	39.4	16.4	2.5	
No Response							
Number	2,096	1,999	6,469	4,719	3,070	581	18,934
Percent at or above	100	88.9	78.4	44.2	19.3	3.1	
Race/Ethnicity							
African American/ Black							
Number	3,438	3,001	7,721	3,518	1,292	111	19,081
Percent at or above	100	82.0	66.3	25.8	7.4	0.6	
American Indian/ Alaskan Native							
Number	237	158	380	186	98	16	1,075
Percent at or above	100	78.0	63.3	27.9	10.6	1.5	
Asian American/ Pacific Islander							
Number	177	170	467	325	225	37	1,401
Percent at or above	100	87.4	75.2	41.9	18.7	2.6	
Caucasian/White							
Number	6,217	6,439	24,774	20,022	12,947	2,338	72,737
Percent at or above	100	91.5	82.6	48.5	21.0	3.2	
Hispanic							
Number	697	546	1,455	758	381	44	3,881
Percent at or above	100	82.0	68.0	30.5	11.0	1.1	
Other, Prefer Not to Respond, No Response							
Number	5,249	4,859	15,414	11,056	6,505	1,121	44,204
Percent at or above	100	88.1	77.1	42.3	17.3	2.5	

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. This requires considerably more work because there would then be many standards with which an individual could be compared. The job profiles being collected by ACT are examples of such multiple standards. For additional information, please consult the *WorkKeys Technical Handbook* (ACT, 2001).

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 an optional 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 Tests of Educational Achievement

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 progress 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 Test

The ACT contains four multiple-choice tests—English, Mathematics, Reading, and Science—and an optional Writing Test. 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, mathematics, reading, science, and writing) 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; and meetings with content experts (see ACT, *ACT National Curriculum Survey® 2005–2006*, 2007a).

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.

The Writing Test (optional)

The ACT Writing Test is a 30-minute essay test that measures students' writing skills—specifically those writing skills emphasized in high school English classes and in entry-level college composition courses. The test consists of one writing prompt that defines an issue and describes two points of view on that issue. The students are asked to respond to a question about their position on the issue described in the writing prompt. In doing so, they may adopt one or the other of the perspectives described in the prompt, or they may present a different point of view on the issue. The essay score is not affected by the point of view taken on the issue.

Taking the Writing Test does **not** affect a student's score on the multiple-choice tests or the Composite score for those tests. Rather, two additional scores are provided: a Combined English/Writing score and a Writing subscore. Also provided are comments on the student's essay.

Test Development Procedures for 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 2.16 through 2.19.

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.

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).

Table 2.16: 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]	015	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 2.17: 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 2.18: 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 2.19: 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.

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 2.16 through 2.19. Items are also selected to comply with the statistical specifications described on pages 36–37. The distributions of item difficulty levels obtained on recent forms of the four tests are displayed in Table 2.20. The data in Table 2.20 are taken from random samples of approximately 2,000 students from each of the six national test dates during the 2005–2006 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.

The 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 2.20 for each test is the average completion rate for the six national test dates during the 2005–2006 academic year. The completion rate for each test is computed as the average proportion of examinees who answered each of the last five items.

Table 2.20: Difficulty[a] distributions and mean discrimination[b] indices for ACT test items, 2000–2001

	Observed Difficulty Distributions (Frequencies)			
	English	**Mathematics**	**Reading**	**Science**
Difficulty Range				
0.00–0.09	0	0	0	0
0.10–0.19	1	8	0	2
0.20–0.29	6	41	9	19
0.30–0.39	17	42	22	32
0.40–0.49	53	66	41	36
0.50–0.59	81	59	42	40
0.60–0.69	101	63	65	49
0.70–0.79	113	49	40	32
0.80–0.89	72	30	21	25
0.90–1.00	6	2	0	5
Number of Items[c]	375	300	200	200
Mean Difficulty	0.64	0.52	0.57	0.55
Mean Discrimination	0.51	0.55	0.51	0.48
Avg. Completion Rate[d]	92	91	93	95

[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.

Content 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 (see Chapter 3 for details about fairness reviews).

Content review panels 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. 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 to ensure all students were tested fairly (see Chapter 3 for details on DIF procedures).

Test Development Procedures for the Writing Test

This section describes the procedures that are used in developing essay prompts for the ACT Writing Test. These include many of the same stages as those used to develop the multiple-choice tests.

Selection and Training of Prompt Writers

ACT holds a prompt writing workshop each year in which new essay prompts are developed. The participants invited to take part in this prompt development process are both high school and post secondary teachers who are specialists in writing, and who represent the diversity of the U.S. population in ethnic background, gender, and geographic location.

Prompt Construction

Prompts developed for the Writing Test provide topics that not only offer adequate complexity and depth so that examinees can write a thoughtful and engaging essay, but also are within the common experiences of high school students. Topics are carefully chosen so that they are neither too vast nor simplistic, and so that they do not require specialized prior knowledge. The topics are designed so that a student should be able to respond to a topic within the 30-minute time constraint of the test.

Content and Fairness Review of Prompts

After Writing Test prompts are developed and then refined by ACT writing specialists, the prompts go through a rigorous review process by external experts. These fairness and bias experts carefully review each prompt to ensure that neither the language nor the content of a prompt will be offensive to a test taker, and that no prompt will disadvantage any student from any geographic, socioeconomic, or cultural background.

Field Testing of Prompts

New Writing Test prompts are field-tested throughout the United States every year. Students from rural and urban settings, small and large schools, and both public and private schools write responses to the new prompts, which are then read and scored by trained ACT readers.

Review of Field Tests and Operational Administration

Once scoring of the new Writing Test prompts has been completed, the prompts are analyzed for acceptability, validity, and accessibility. The new field-tested prompts are also reviewed to ensure that they are compatible with previous operational prompts, that they function in the same way as previous prompts, and that they adhere to ACT's rigorous standards.

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.

In addition to the above scores, if the student took the Writing Test, the student's essay is read and scored independently by two trained readers using a six-point scoring rubric. Essays are evaluated on the evidence they demonstrate of student ability to make and articulate judgments; develop and sustain a position on an issue; organize and present ideas in a logical way; and communicate clearly and effectively using the conventions of standard written English. Essays are scored holistically—that is, on the basis of the overall impression created by all the elements of the writing. Each reader rates an essay on a scale ranging from 1 to 6. The sum of the readers' ratings is a student's Writing Test subscore on a scale ranging from 2 to 12. A student who takes the Writing Test also receives a Combined English/Writing score on a score scale ranging from 1 to 36. Writing Test results do not affect a student's Composite score.

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.

If a student believes that a Writing Test essay has been incorrectly scored, that score may be appealed, and the essay will be reviewed and rescored (for a fee) by two new expert readers. The two new readers score the appealed essay without knowledge of the original score, and the new score is adjudicated by ACT staff writing specialists before being finalized.

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 (2007). The ACT Technical Manual (2007) can be acquired from ACT's website at **www.act.org**.

The Alignment of PSAE Scores with the Illinois Learning Standards

Each of the PSAE content scores is based on the combination of two sets of items, one from Day 1 and one from Day 2. The scaling process is described in detail in Chapter 4, but briefly, the Day 1 and Day 2 component scores are equally weighted to form an overall PSAE assessment score.

The decision to equally weight the components was based on a number of considerations. The primary consideration was the alignment between the Illinois Learning Standards and the items contained within the component assessments. All components of the PSAE were reviewed and evaluated in terms of their overall alignment to the Illinois Learning Standards. Based on several independent reviews, all items on every PSAE component were found to be aligned to the Illinois Learning Standards.

Secondary considerations for equal weights included providing an incentive for students to be equally motivated to try their best on both days, thus providing a more accurate view of student achievement. Technical characteristics of the Day 1 and Day 2 components, such as the number of items, difficulty of items, and component reliability were also reviewed prior to finalizing the equal weighting decision.

The 90 PSAE mathematics items all align to the Illinois Learning Standards in mathematics, and cover the Illinois Learning Standards well in terms of content, skills, context (covering both academic and workplace contexts as required by the "Applications of Learning" sections within the Illinois Learning Standards in Mathematics), and range of acquired abilities.

The 70 PSAE reading items all align to the Illinois Learning Standards in reading, and cover the Illinois Learning Standards well in terms of skills, context (covering both academic and workplace contexts as required by the "Applications of Learning" sections within the Illinois Learning Standards in reading), and range of acquired abilities.

The 80 PSAE science items all align to the Illinois Learning Standards in Science, and cover the Illinois Learning Standards well in terms of skills, context, and range of acquired abilities across the content areas of biology, chemistry, physics, and earth/space science.

The items across the two assessment components for each PSAE area provide the desired coverage of the Illinois Learning Standards as defined by the PSAE Assessment Frameworks shown in Tables 2.21, 2.24, and 2.27.

When looking at the effective percents of the coverage of the PSAE items, which take into account the weighting factor for the PSAE score, to the Assessment Frameworks, the effective percents are similar to or match exactly the PSAE Framework requirements and also similar to or match exactly the optimal percents given the weighting factor. Tables 2.21 through 2.29 below show the assessment frameworks, effective percents, and optimal percents in the areas of Mathematics, Reading, and Science.

Table 2.21: PSAE Mathematics Assessment Framework

Mathematics Illinois State Goals	PSAE Framework Approx. %	Day 1 Pool	Day 2 Pool
Goal 6 – Number Sense	33%	18%	63%
Goal 7 – Measurement	20%	11%	37%
Goal 8 – Algebra	27%	40%	0%
Goal 9 – Geometry	16%	24%	0%
Goal 10 – Data/Stat/Prob	4%	7%	0%

Table 2.22: Effective item counts and percents for Mathematics given 50/50 weighting

Mathematics Illinois State Goals	Effective Item Counts			Effective Percent
	PSAE	Day 1 Pool	Day 2 Pool	
Goal 6 – Number Sense	24	5	19	40%
Goal 7 – Measurement	14	3	11	23%
Goal 8 – Algebra	12	12	0	20%
Goal 9 – Geometry	7	7	0	12%
Goal 10 – Data/Stat/Prob	2	2	0	4%
	60	30	30	

Table 2.23: Optimal counts and percents for Mathematics

Mathematics Illinois State Goals	Items Needed	Optimal Item Counts		Score Points	Optimal %
		Day 1	Day 2		
Goal 6 – Number Sense	24	5	19	21	35%
Goal 7 – Measurement	14	3	11	12	20%
Goal 8 – Algebra	12	12	0	16	27%
Goal 9 – Geometry	7	7	0	9	15%
Goal 10 – Data/Stat/Prob	2	2	0	2	3%
	60	Day 2 + ½(Day 1) = 60			

Table 2.24: PSAE Reading Assessment Framework

Reading Illinois State Goals	PSAE Framework Approx. %	Day 1 Pool	Day 2 Pool
Goal 1 – Reading	100%	100%	100%

Table 2.25: Effective item counts and percents for Reading given 50/50 weighting

Reading Illinois State Goals	Effective Item Counts			Effective Percent
	PSAE	Day 1 Pool	Day 2 Pool	
Goal 1 – Reading	60	30	30	100%
	60	30	30	

Table 2.26: Optimal counts and percents for Reading

Reading Illinois State Goals	Items Needed	Optimal Item Counts		Score Points	Optimal %
		Day 1	Day 2		
Goal 1 – Reading	60	30	30	60	100%
	60	Day 2 + ¾(Day 1) = 60			

Table 2.27: PSAE Science Assessment Framework

Science Illinois State Goals	PSAE Framework Approx. %	Day 1 Pool	Day 2 Pool
Goal 11 – Science Inquiry	52%	100%	5%
Goal 12 – Science Content	38%		75%
Goal 13 – Sci/Tech/Society	10%		20%

Table 2.28: Effective item counts and percents for Science given 50/50 weighting

Science Illinois State Goals	Effective Item Counts			Effective Percent
	PSAE	Day 1 Pool	Day 2 Pool	
Goal 11 – Science Inquiry	42	40	2	52%
Goal 12 – Science Content	30	0	30	38%
Goal 13 – Sci/Tech/Society	8	0	8	10%
	80	40	40	

Table 2.29: Optimal counts and percents for Science

Science Illinois State Goals	Items Needed	Optimal Item Counts		Score Points	Optimal %
		Day 1	Day 2		
Goal 11 – Science Inquiry	42	40	2	42	52%
Goal 12 – Science Content	30	0	30	30	38%
Goal 13 – Sci/Tech/Society	8	0	8	8	10%
	80	Day 1 + Day 2 = 80			

As can be seen in Tables 2.21 through 2.29, the optimal percents, match the effective percents and the framework percents exactly for Reading and Science. For Mathematics, the optimal percents are similar to the effective percents and the framework percents but do not match exactly. PSAE documents mention ranges of scores and items within the PSAE Mathematics Assessment Frameworks. Listing ranges of scored items within goals in the PSAE assessment frameworks better reflects the test development process and test blueprint in mathematics especially, given the large number of skills and knowledge within each goal. ISBE in concert with stakeholder groups will reexamine the assigned values within the PSAE Assessment Frameworks and develop ranges that reflect the rigor and dynamics of Illinois Learning Standards.

Along with evaluating the effective percents against the Assessment Framework percent requirements and optimal percents, principal component analysis of the correlation matrix for the three PSAE areas is conducted each year to ensure that there is consistency of the PSAE construct across years. This analysis has shown that for all three PSAE areas (i.e., Reading, Math, and Science) the first principal component loadings are very high and nearly equal across years. For Reading, the first principal component loading values range from .91 to .93 across all six administrations of the PSAE (i.e., from 2001 to 2006). For Math, the first principal component loading values range from .91 to .94 across all six years. And for Science, the first principal component loading values range from .94 to .96 across all six years. The first factor component's variance for all years (2001 through 2006) was equal to or greater than 83%.

Chapter 3
PSAE Sensitivity and Bias Procedures and Analyses of DIF Results

Commitment to Fairness

The purposes of this chapter are to describe the sensitivity and bias procedures followed during development of the PSAE that ensure that these tests are as fair as possible to all examinees who take them, and to present results of analyses documenting that the PSAE test forms 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 Illinois state goals and standards. If factors other than the academic skills and knowledge in those content areas intrude, 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 both in principle and in the interest of accuracy in 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 pretesting, item selection and forms construction, and forms review. Every effort is made to see that PSAE test forms are fair for Illinois students. This chapter will begin with describing those procedures, and end with tabulated results.

Fairness and Bias Reviews

Preliminary versions of PSAE items, Writing Test prompts, and test forms are subjected to several reviews to ensure that the test materials are fair and conform to good test construction practice. The first fairness review is performed by PSAE test development staff at ACT. Items are 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 items are then submitted to 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). Writing Test prompts also go through a rigorous review process by external experts. These fairness and bias experts carefully review each prompt to ensure that neither the language nor the content of a prompt will be offensive to a test taker, and that no prompt will disadvantage any student from any geographic, socioeconomic, or cultural background.

Fairness review panels are convened to discuss with PSAE development staff the consultants’ reviews of the PSAE test forms. Fairness review panels consist of content experts, and experts in diverse areas of education who represent both genders and a variety of racial and ethnic backgrounds. The fairness panels review the forms to help ensure fairness to all PSAE examinees.

After the panels complete their reviews, comments from the consultants are reviewed by PSAE test developers 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 PSAE test developers. If no further corrections are needed, the test forms are prepared for printing.

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.

A summary of the DIF analysis results for the PSAE Standard form administered in Spring 2006 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 Science only). The table indicates that in Mathematics, for example, 2 out of the 90 items administered on the standard form appeared to favor males, and 2 appeared to favor females, based on the statistical indices. A total of 17 out of the 717 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.

Table 3.1: Summary of DIF Analysis Results for the PSAE Standard Form Administered in Spring 2006

Favored Group	Subject		
	Reading	Mathematics	Science
Male		2	4
Female		2	
African American		1	
Caucasian	3	2	1
Hispanic American			
Caucasian	1		1

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

Scale scores are reported for PSAE mathematics, reading, and science. All three of these scales are based on combinations of two assessments. The following descriptions pertain to the PSAE mathematics, reading, 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 five tests. The means and standard deviations pertain to grade 11 students in Illinois public schools.

Scaling of the PSAE Mathematics, Reading, 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 scales. This section contains a discussion of the data used in scaling the PSAE assessment.

The Scaling Process

Raw scores for the PSAE composites were created by combining the scores of the two components, using equal weighting. Number-of-items-correct scores on each of the two components were transformed to equated raw scores for the underlying scale. For example, raw scores on the ACT Mathematics Test were transformed to equated raw scores on an underlying ACT Mathematics Test raw score scale, using a base form. A similar procedure was followed for the other ACT tests and the two WorkKeys assessments. For the ISBE science test, the raw scores of the form administered in April were used as the base form. The equated or base-form raw scores, as appropriate, of the two components of each test were converted to the standard normal scores (Z-scores) using the means and standard deviations of the equated or base-form raw scores from the scaling group. Then, the Z-scores on each component were combined with the weight of 0.5 for each component. The equally weighted composite Z-scores were referred to as the PSAE raw scores.

The scaling process for the PSAE scales consisted of four steps. First, PSAE raw score distributions for the scaling group were computed. In the second step, the PSAE raw scores were linearly or nonlinearly, depending on the PSAE raw score distributions, transformed to produce initial scale scores with the targeted means and standard deviations. The distributions for the PSAE raw scores for reading and science were close to normal, so they were linearly transformed to produce the initial scale scores. The raw score distribution for the mathematics test was positively skewed, which could result in a maximum initial scale score substantially smaller than the targeted maximum scale score. Therefore, the PSAE raw scores for mathematics were normalized and then linearly transformed to produce initial scale scores. In the third step, the initial score scales on the tests were rounded to integers and truncated to range from 120 to 200.

Finally, some of the rounded scale scores were adjusted to attempt to meet the specified targets for means and standard deviations, to avoid gaps in the score scale, or to avoid having too many raw scores converting to a single scale score. This scaling process resulted in the final raw–to–scale-score conversions. It should be noted that the final conversions are based on, for example, a raw ACT Reading Test score and a raw WorkKeys reading score. These two raw scores are transformed to equated raw scores, which are then converted to Z-scores; the Z-scores are then combined using equal weighting to form a PSAE raw score, which is then transformed to a PSAE scale score using the PSAE raw–to–scale-score conversion.

The raw-to-scale-score transformations of the PSAE assessment components are presented in Figure 4.1. The raw-to-scale-score transformations for the PSAE reading and science scales are linear in the middle part of the scale-score ranges, in contrast to the transformations in both tails, which are not linear because of truncation or score adjustments. For example, many PSAE raw scores in the lower tails of the PSAE reading scale are truncated to the scale score of 120. The raw–to–scale-score transformation of the PSAE science scale has a steeper slope in the scale-score range of between 190 and 200, compared to the other scale-score ranges. The raw–to–scale-score transformation of the PSAE mathematics scale departs considerably from linearity, compared to the other PSAE scales, and is flat at the extremely low raw scores because of truncations.

Summary Statistics

Scale-score summary statistics for the scaling 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 160 and 15, respectively, which were the targeted mean and standard deviation. Because of rounding, truncations, and some score adjustments, the means and standard deviations in Table 4.1 were not exactly the same as the targeted ones.

Figure 4.1: Raw-to-scale-score transformation for the PSAE tests

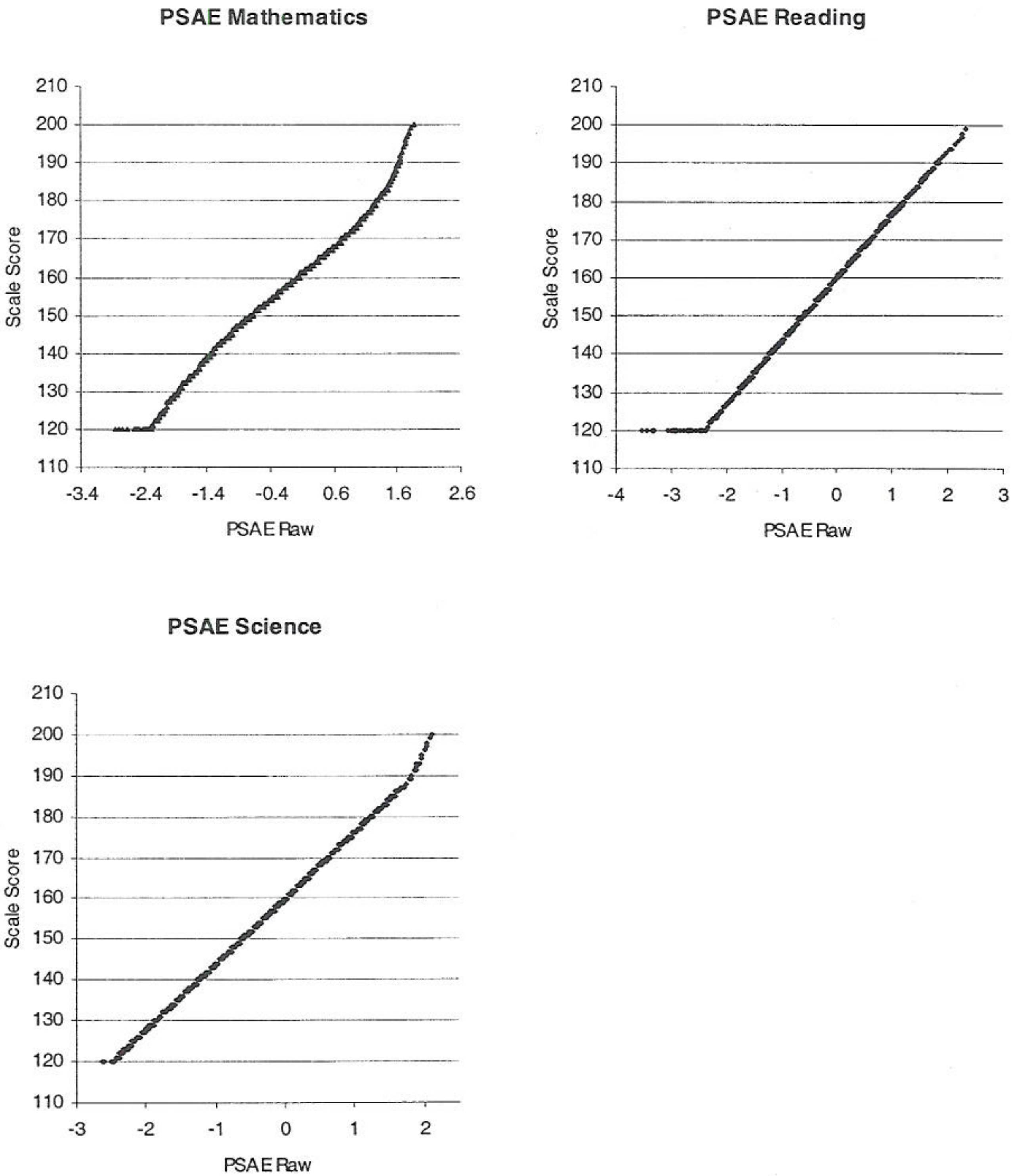


Table 4.1: Scale-score summary statistics for the PSAE scales for the scaling group (N = 10,554)

Statistics	Mathematics	Reading	Science
Mean	159.9845	160.0183	160.0232
SD	15.0095	14.9342	15.3464
Skewness	0.0114	–0.1013	–0.1743
Kurtosis	–1.1819	–0.4486	–0.6761

Measurement Error and Reliability for the PSAE Scores

The conditional standard errors of measurement summarize the amount of error or inconsistency of reported scores at different points on the score scale. The binomial error model (Lord, 1955, 1957) has been widely known as a way of computing conditional standard errors of measurement for raw scores when a test consists of dichotomously scored items. Kolen, Hanson, and Brennan (1992) provide a general framework for obtaining conditional standard errors of measurement for scale scores. To compute the conditional standard errors of measurement on the PSAE scales, the procedure of Kolen and others needs to be extended because the three PSAE assessment areas consist of two components.

Because the components of the PSAE mathematics, reading, and science assessments contain only dichotomously scored items and there are two components for the assessments, assuming the errors of each component are independent, the binomial error model can be extended to compute the conditional standard errors of measurement of scale scores. The specific procedures are as follows:

1. The observed proportion-correct scores (the number of items correct divided by the total number of items) of the examinees on each component of a particular PSAE assessment are treated as the true proportion-correct scores,
2. using the proportion-correct scores as binomial probabilities, the binomial distributions are computed for each component,
3. given two proportion-correct scores on two components and assuming the number correct scores given true scores are independent for the two components, the bivariate binomial probability density functions (BBPDF) are obtained,
4. using the raw–to–scale-score conversion and BBPDF, the conditional mean and standard errors of measurement are computed, and
5. the squared conditional standard errors of measurement are averaged across the students, which results in the average conditional error variance that will be used for computing reliability for scale scores.

Measurement Error and Reliability for the PSAE Scores for the Scaling Group

The conditional standard errors of measurement for the mathematics, reading, and science scales are presented in Figures 4.2 through 4.7. The data used to produce these figures are the same samples used for scaling the components of each PSAE test. The conditional standard errors of measurement in Figures 4.2, 4.4, and 4.6 are plotted as a function of the observed scale scores across each PSAE subject test. Different conditional standard errors of measurement for each particular value of observed scale score are possible because more than one combination of scores for the two components of each PSAE test can produce the same scale score. For example, if some students receive the scale score 175 for PSAE mathematics, the standard error of measurement of the score can range from about 2.5 to 4.5, depending on the different ways students performed on the two components of the PSAE mathematics test, the ACT Mathematics Test and WorkKeys *Applied Mathematics*. The conditional standard errors of measurement, for scale scores, in Figures 4.3, 4.5, and 4.7 are plotted three-dimensionally as a function of both the observed number-correct scores of the two PSAE test components. The point on the surface of the plane that intersects with the raw scores of the two components of the PSAE subject test indicates the conditional standard errors of measurement for the scale scores.

With the binomial error model assumption, the conditional standard errors of measurement are large at the middle part and small at the tails of the raw scores because of the characteristics of the binomial model. The shape of the conditional standard errors of measurements for the raw scores typically looks like an inverted U-shape across the score ranges. When there are two components, as is the case for the PSAE tests, the shape of the conditional standard errors of measurement for the raw scores is analogous to the inverted U-shape. However, the shape of the conditional standard errors of measurements for scale scores are affected by the raw–to–scale-score transformation employed (Brennan & Lee, 1997, p.13; Kolen, Hanson & Brennan, 1992). It is also true for PSAE scale scores. For example, when the raw–to–scale-score transformations are linear (or close to linear) as that of PSAE reading (see the top right of Figure 4.1), the shapes of the conditional standard errors of measurement are similar to the analogue of the inverted U-shape (see Figures 4.4 and 4.5). In contrast, when the raw–to–scale-score transformations depart from linearity as that of PSAE mathematics (see the top left of Figure 4.1) or as that of the upper range of PSAE science scores (see the bottom left of Figure 4.1), the conditional standard errors of measurement display irregular patterns (here a bimodal-like shape).

The conditional standard errors of measurement are used to construct the confidence interval for the examinees' true scale scores. That is, by adding and subtracting one conditional standard error of measurement from examinees' observed scale scores, the approximate 68% of confidence intervals for their true scores are constructed.

The estimated scale-score reliability for the PSAE assessment i (rel_i), where i = the PSAE mathematics, reading, and science assessments, is calculated as

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

where $\overline{\sigma}^2(E_i)$ is the average of the estimated conditional error variance and $\sigma^2(S_i)$ is the observed scale-score variance for test i. Table 4.2 shows the average standard errors of measurement and reliabilities for the PSAE tests. The variation in the standard errors of measurement among the PSAE tests occurred as a result of differences in raw score means, reliabilities, and test lengths on each component constituting the PSAE, which leads to different reliabilities among the components.

Table 4.2: Average standard errors of measurement (SEMs) and reliabilities for the scaling group (N = 10,554)

Statistics	Mathematics	Reading	Science
SEM	4.0328	6.7458	4.5545
Reliability	0.9278	0.7960	0.9119

Figure 4.2: PSAE Mathematics – Conditional standard errors of measurement (CSEM) by observed scale score for scaling group

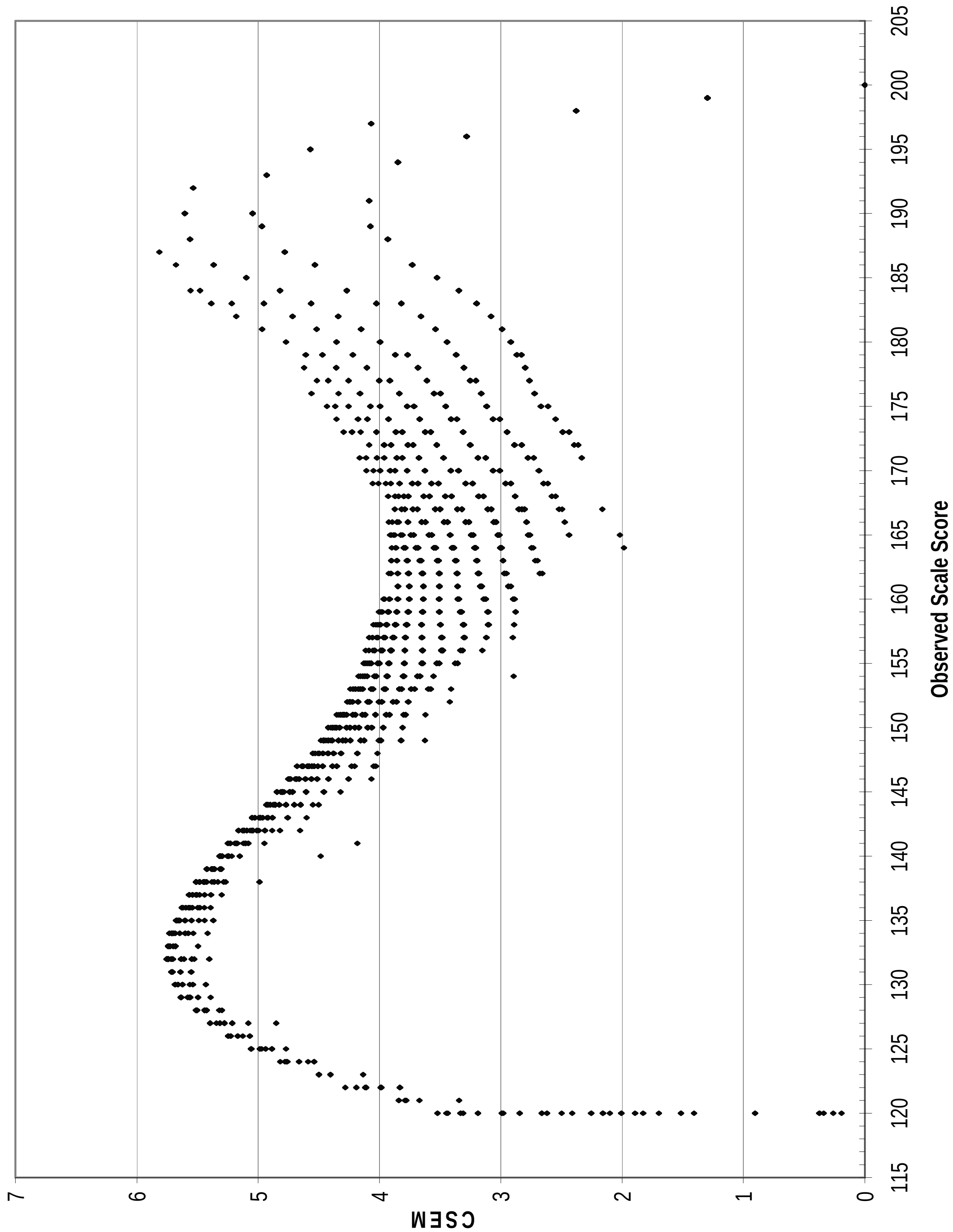


Figure 4.3: PSAE Mathematics – Conditional standard errors of measurement (CSEM) by component raw scores for scaling group

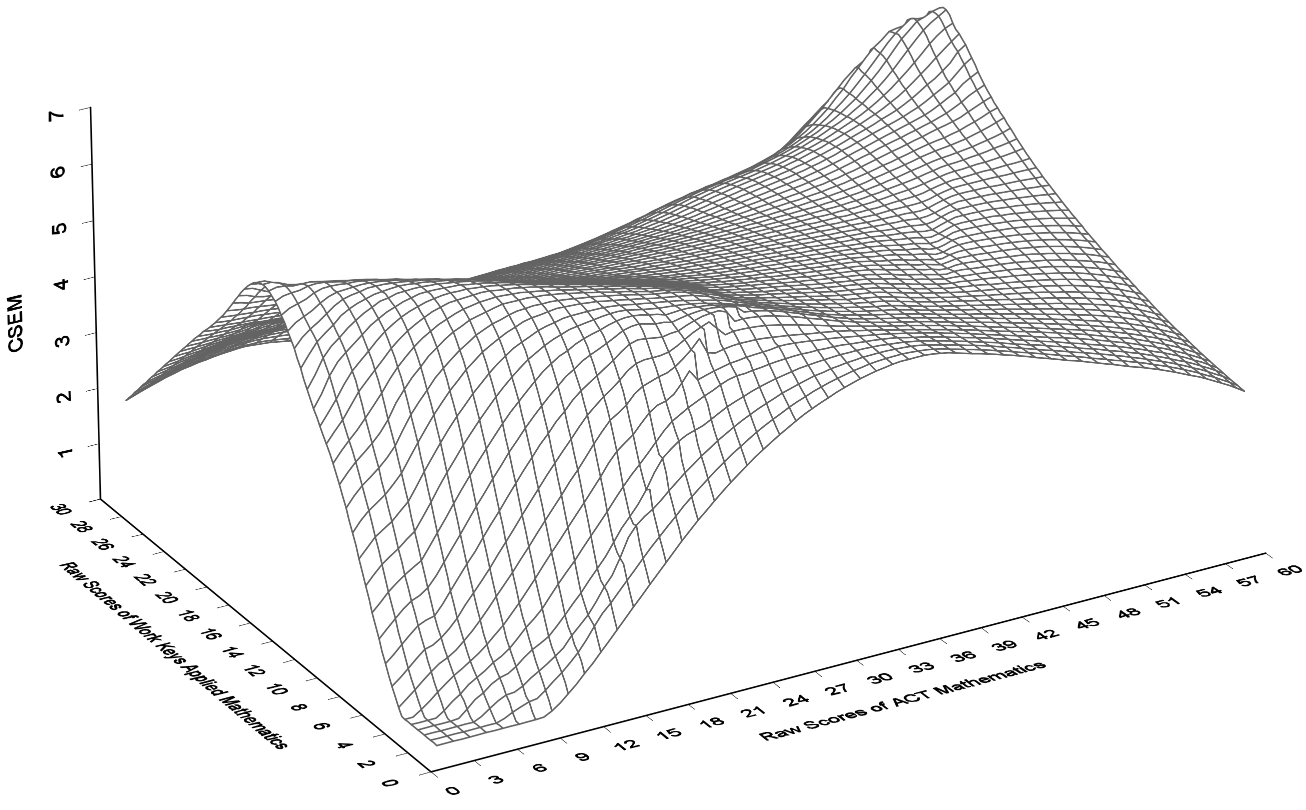


Figure 4.4: PSAE Reading – Conditional standard errors of measurement (CSEM) by observed scale score for scaling group

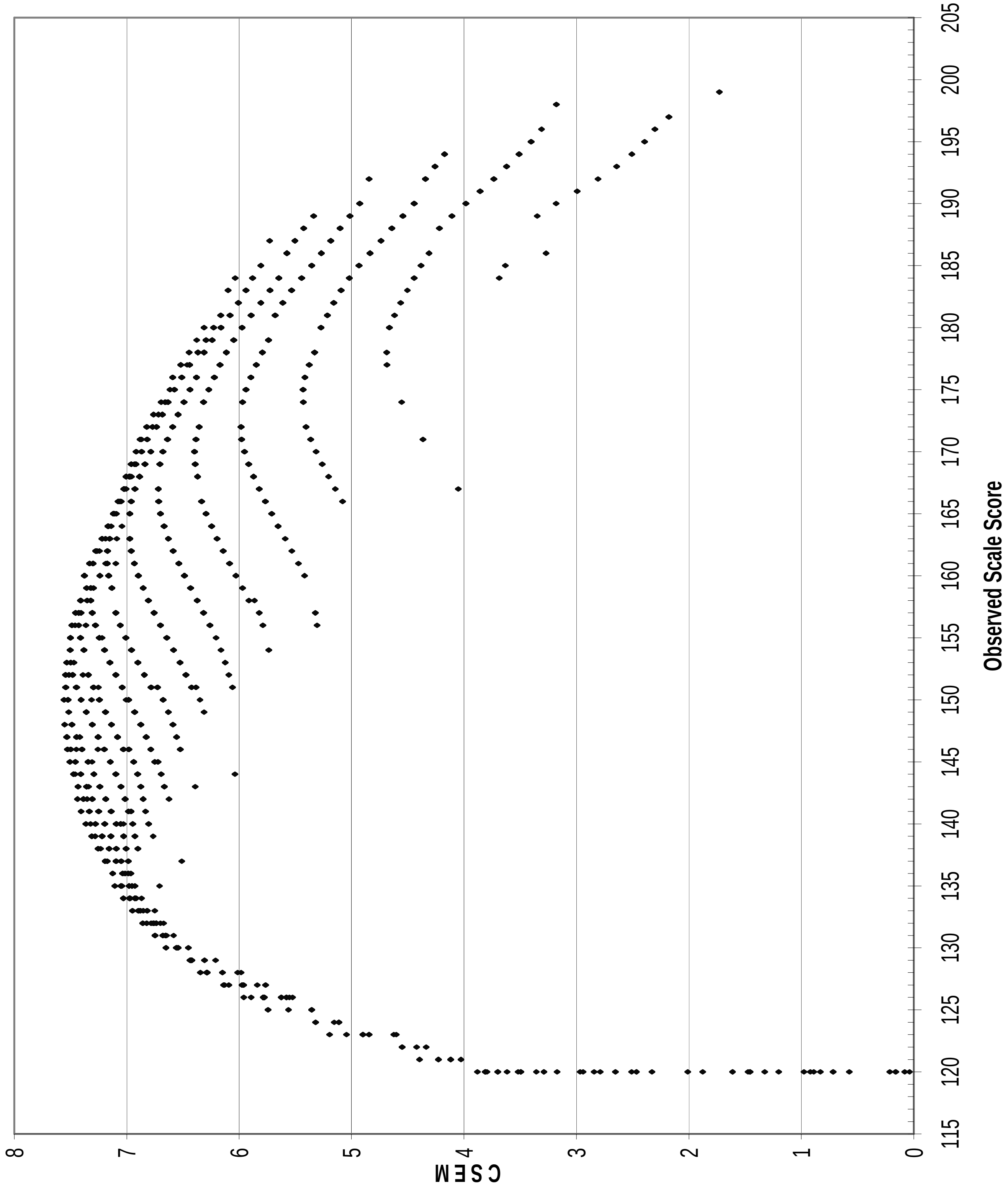


Figure 4.5: PSAE Reading – Conditional standard errors of measurement (CSEM) by component raw scores for scaling group

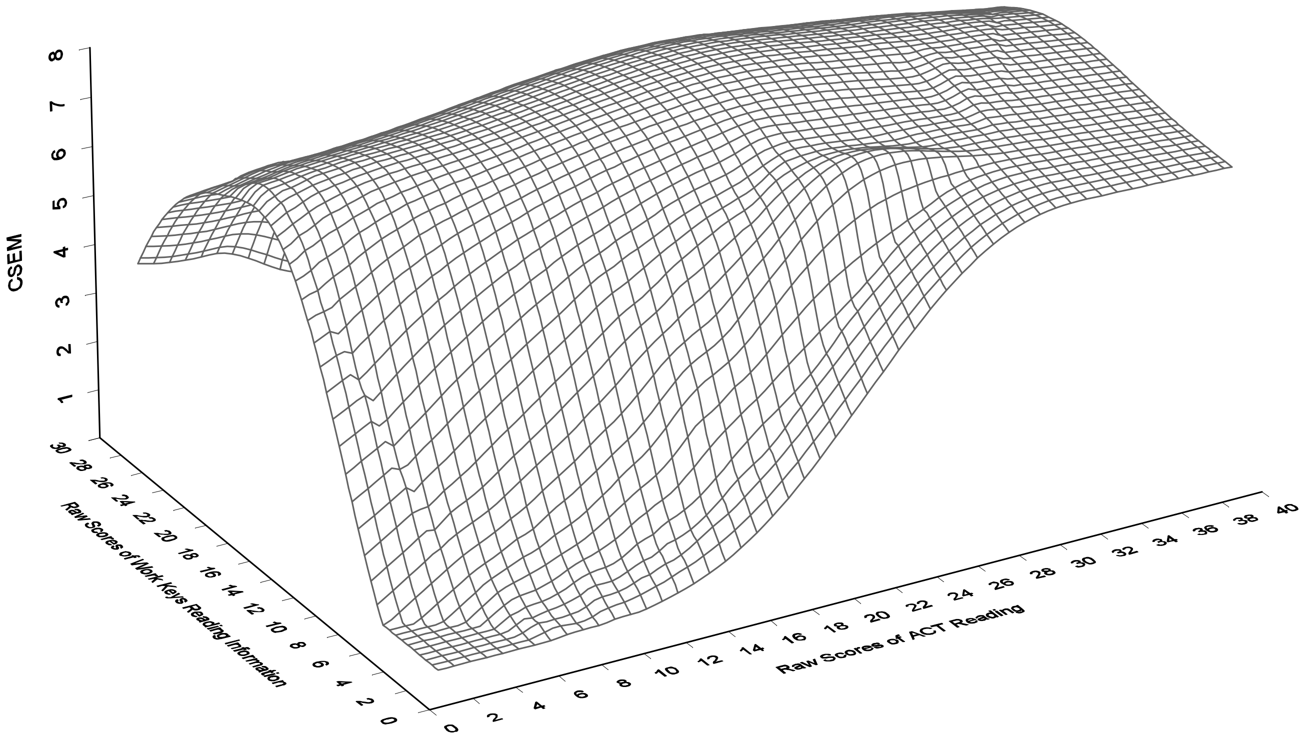


Figure 4.6: PSAE Science – Conditional standard errors of measurement (CSEM) by observed scale score for scaling group

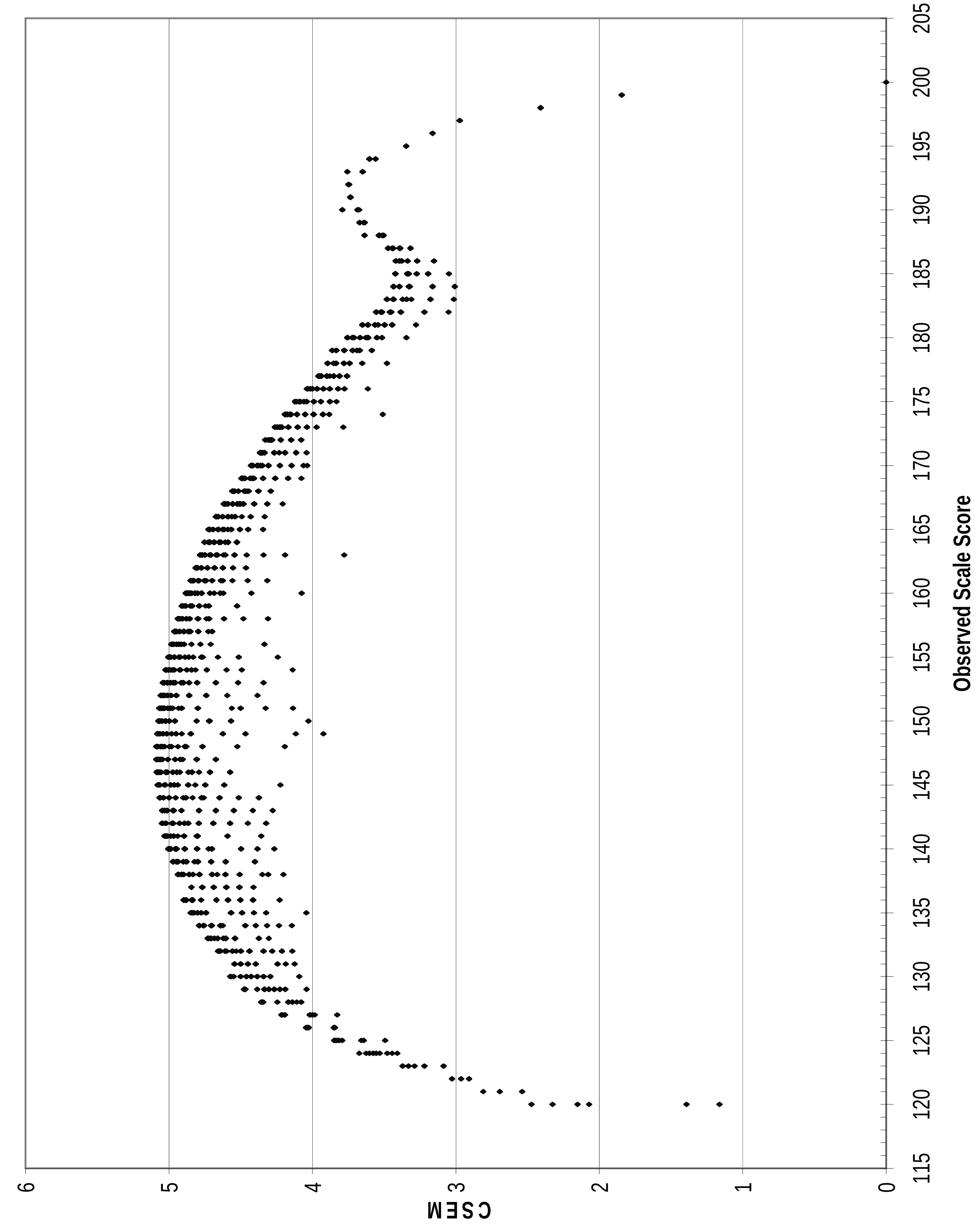
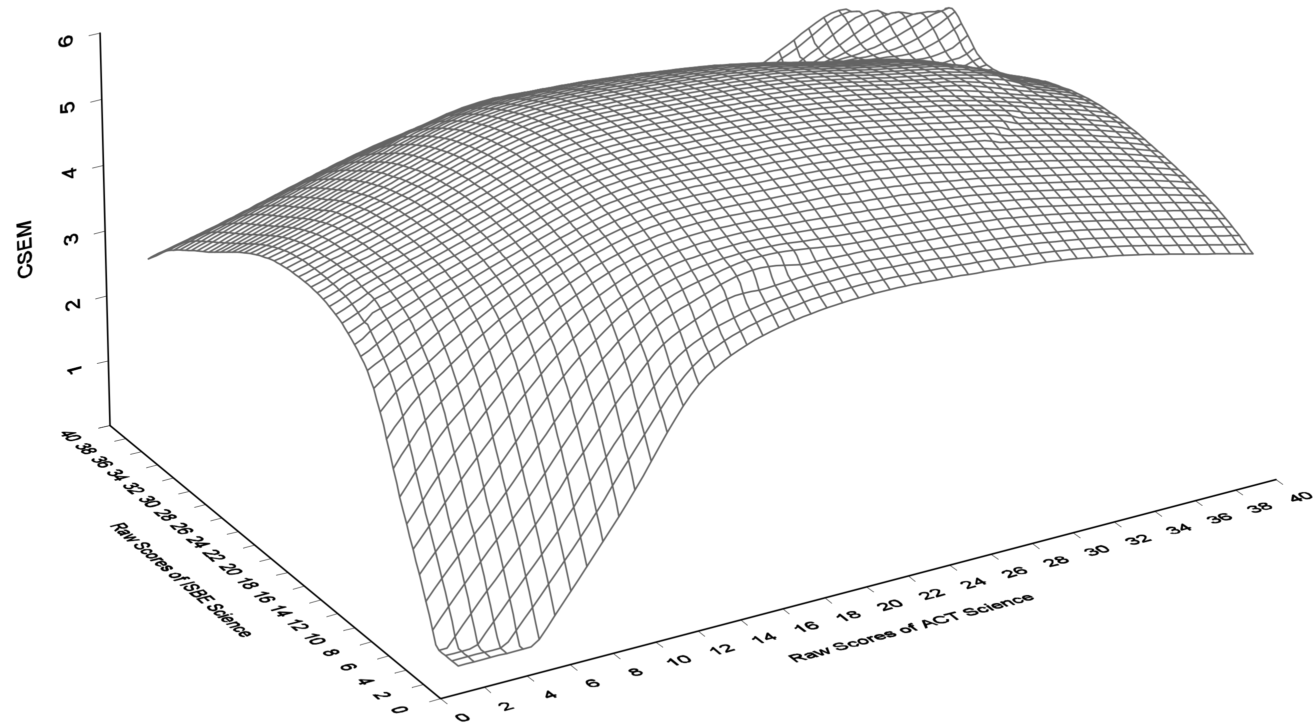


Figure 4.7: PSAE Science – Conditional standard errors of measurement (CSEM) by component raw scores for scaling group



Measurement Error and Reliability for the PSAE Scores for the Spring 2006 Administration

The conditional standard errors of measurement for the mathematics, reading, and science scales for the PSAE Spring 2006 administration are presented in Figures 4.8 through 4.10.

Table 4.3 shows the average standard errors of measurement and reliabilities for the PSAE 2006 spring administration. Examinees with a valid scale score were used in the analyses. The sample sizes differ across content areas because not all examinees had valid scores on all tests.

In general, the 2006 SEM and reliabilities are close to those based on the scaling data. One exception is a higher reliability in reading.

Figures 4.8 through 4.10 illustrate the conditional standard errors of measurement in mathematics, reading, and science in 2006. These plots are very similar to those based on the scaling data and demonstrate a high degree of stability over time.

Table 4.3: Average standard errors of measurement (SEMs) and reliabilities for the PSAE spring 2006 administration

Statistics	Mathematics	Reading	Science
SEM	4.25497	6.64238	4.69418
Reliability	0.91800	0.852589	0.898896
N	111,993	111,977	111,980

Figure 4.8: PSAE Mathematics – Conditional standard errors of measurement (CSEM) by observed scale score for the PSAE Spring 2006 administration

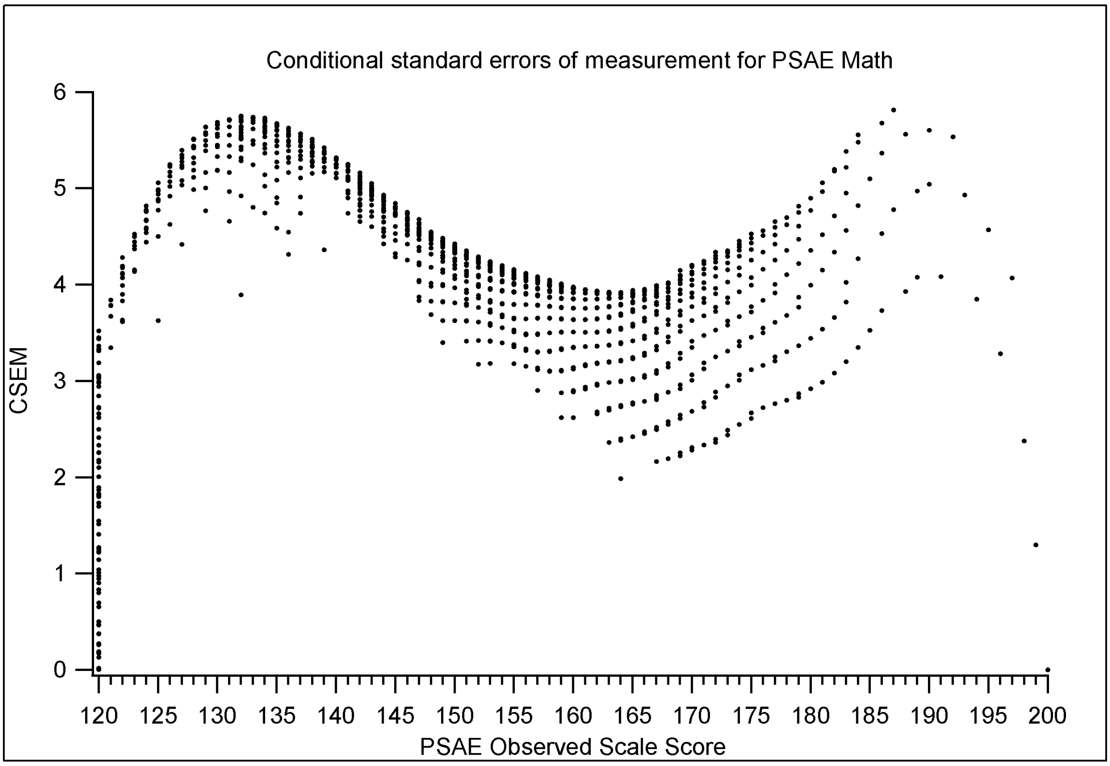


Figure 4.9: PSAE Reading – Conditional standard errors of measurement (CSEM) by observed scale score for the PSAE Spring 2006 administration

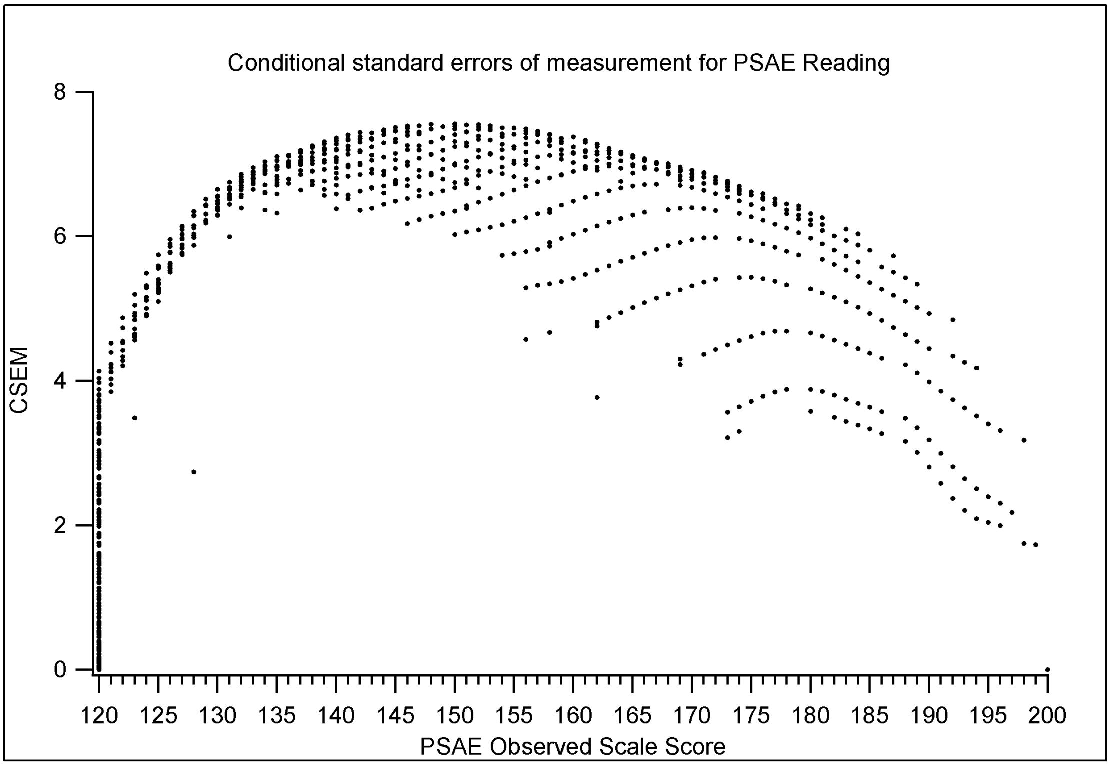
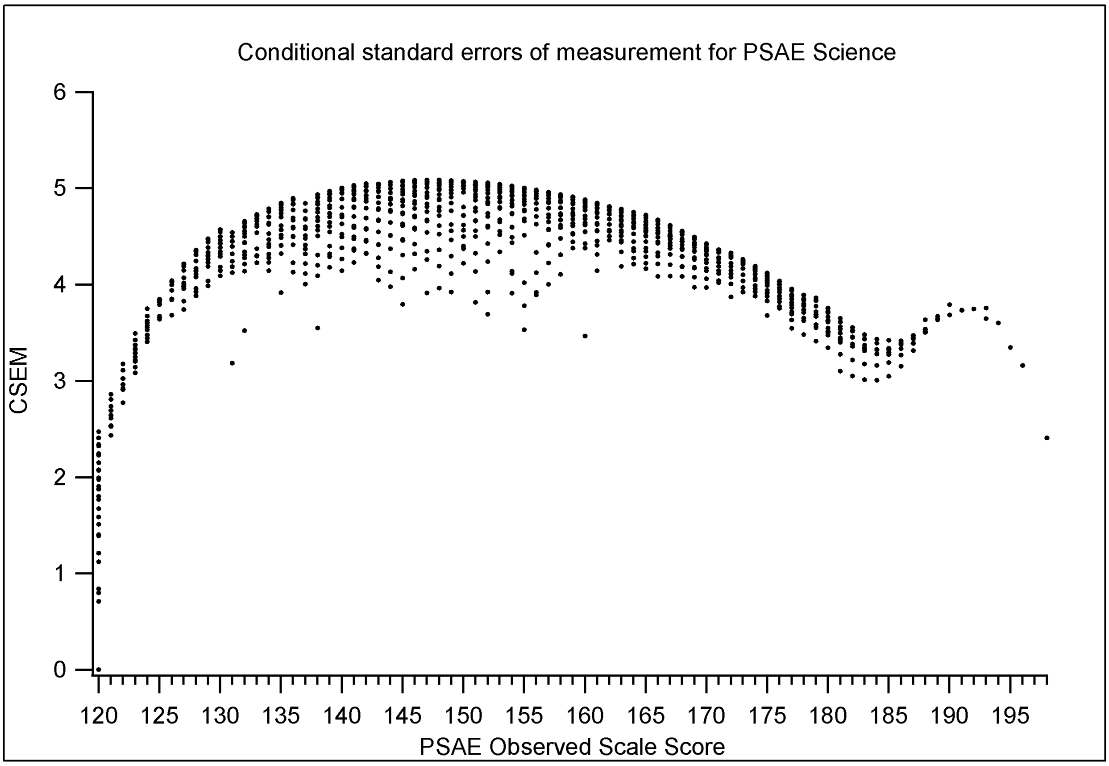


Figure 4.10: PSAE Science – Conditional standard errors of measurement (CSEM) by observed scale score for the PSAE Spring 2006 administration



Chapter 5
Classification Consistency for the PSAE

Classification consistency indicates how consistently examinees are assigned into one of several categories over two independent administrations of a test. 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) are not often administered twice to the same sample.

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. A description of the standard-setting process can be found in Chapter 4 of the 2005 PSAE Technical Manual. The three cut points are different for each of the three PSAE tests (see Table 4.14 in the 2005 PSAE Technical Manual for final cut points). Since there are four categories, 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 the same samples used for scaling the PSAE tests. The procedures for estimating classification consistency are the same up to the steps getting bivariate binomial probability density functions (BBPDF) as those for producing the conditional standard errors of measurement. Once the raw–to–scale-score conversion, BBPDF, and the cut points for scale scores are available, the probabilities being classified in each pair of categories (4 × 4) are computed. These probabilities are summed and divided by the total number of the examinees. By summing the probabilities in the diagonal elements in the 4 × 4 tables, classification consistencies are estimated. Table 5.1 shows 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.

The classification consistency data represented in Table 5.2 are based on the Spring 2006 PSAE data and are the same as those used in the reliability analyses for which results are shown in Table 4.3 of Chapter 4 in this manual. The joint probabilities of being classified in the cells of the 4 × 4 contingency tables are similar to those observed in Table 5.1. Further, the three classification indices in Table 5.1 round to within .02 of those in Table 5.1. The similarities between the values in Tables 5.1 and 5.2 indicate a high degree of stability in the PSAE measurement procedure.

Table 5.1: Classification consistency for the PSAE 2001 scaling data

PSAE Mathematics				
	Academic Warning	**Below**	**Meets**	**Exceeds**
Academic Warning	0.0407	0.0236	0.0000	0.0000
Below	0.0236	0.2435	0.0524	0.0000
Meets	0.0000	0.0524	0.4187	0.0249
Exceeds	0.0000	0.0000	0.0249	0.0953
Classification Consistency:	0.7982			

PSAE Reading				
	Academic Warning	**Below**	**Meets**	**Exceeds**
Academic Warning	0.0342	0.0281	0.0016	0.0000
Below	0.0281	0.1863	0.0878	0.0006
Meets	0.0016	0.0878	0.3337	0.0532
Exceeds	0.0000	0.0006	0.0532	0.1032
Classification Consistency:	0.6574			

PSAE Science				
	Academic Warning	**Below**	**Meets**	**Exceeds**
Academic Warning	0.0512	0.0259	0.0000	0.0000
Below	0.0259	0.2635	0.0596	0.0000
Meets	0.0000	0.0596	0.3286	0.0369
Exceeds	0.0000	0.0000	0.0369	0.1117
Classification Consistency:	0.7550			

Table 5.2: Classification consistency for the PSAE Spring 2006 administration

PSAE Mathematics				
	Academic Warning	**Below**	**Meets**	**Exceeds**
Academic Warning	0.0569	0.0349	0.0000	0.0000
Below	0.0349	0.2930	0.0542	0.0000
Meets	0.0000	0.0542	0.3663	0.0202
Exceeds	0.0000	0.0000	0.0202	0.0651
Classification Consistency:	0.7812			

PSAE Reading				
	Academic Warning	**Below**	**Meets**	**Exceeds**
Academic Warning	0.0563	0.0373	0.0017	0.0000
Below	0.0373	0.1911	0.0754	0.0005
Meets	0.0017	0.0754	0.2829	0.0508
Exceeds	0.0000	0.0005	0.0508	0.1381
Classification Consistency:	0.6685			

PSAE Science				
	Academic Warning	**Below**	**Meets**	**Exceeds**
Academic Warning	0.0789	0.0447	0.0000	0.0000
Below	0.0447	0.3609	0.0595	0.0000
Meets	0.0000	0.0595	0.2525	0.0229
Exceeds	0.0000	0.0000	0.0229	0.0533
Classification Consistency:	0.7457			

Chapter 6
Ensuring Consistency of PSAE Score Meaning Over Time

The PSAE program is administered in April, with a makeup administration in May; a retake for grade 12 students is offered in October. So that scores from these different administrations are comparable, as well as to allow tracking of trends across time, new forms of the ACT Test, 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. These equated raw scores are then converted to Z-scores, combined using equal weighting, and converted using the PSAE raw–to–scale-score conversion.

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; this process retains the equal weighting of components across PSAE forms, and allows PSAE scores from 2006 to be compared to PSAE scores from 2001.

Equating of ISBE forms

New forms of the ISBE 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 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 each of the ACT Assessment tests 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.

As specified in the *Standards for Educational and Psychological Testing* (APA, 1985), ACT conducts periodic checks on the stability of the ACT Assessment 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.)

Tables 6.1 through 6.6 provide information relating PSAE scores in math, reading, and science, respectively, to the component scale scores. For all students with valid PSAE scores, the conditional mean PSAE score was calculated for all students with a given ISBE Science scale score, ACT scale score, or WorkKeys level score. In Table 6.1, it can be seen that, for example, an ACT score of 30 is associated with a PSAE score of 181 in 2006, and a score of 182 in 2001, a difference of only one PSAE score point. Difference are small throughout the score range, which is also true for the rest of the tables, indicating that the PSAE scores appear stable using these criteria.

Table 6.1: Relationship of PSAE Mathematics Scores to ACT Mathematics Scale Scores

ACT Mathematics	PSAE Mathematics 2006	PSAE Mathematics 2001	Diff (2006 – 2001)
1	121	127	–6
2	NA	NA	NA
3	NA	122	NA
4	122	NA	NA
5	NA	123	NA
6	123	127	–4
7	NA	124	NA
8	126	121	5
9	122	124	–2
10	125	126	–1
11	127	128	–1
12	130	132	–2
13	132	134	–2
14	136	138	–2
15	141	142	–1
16	147	148	–1
17	152	152	0
18	155	155	0
19	158	158	0
20	161	161	0
21	163	162	1
22	164	164	0
23	166	166	0
24	168	168	0
25	170	170	0
26	172	173	–1
27	175	175	0
28	177	177	0
29	179	180	–1
30	181	182	–1
31	184	184	0
32	186	188	–2
33	188	191	–3
34	191	194	–3
35	194	196	–2
36	197	198	–1

Table 6.2: Relationship of PSAE Mathematics Scores to WorkKeys Applied Mathematics Level Scores

WK Mathematics	PSAE Mathematics 2006	PSAE Mathematics 2001	Diff (2006 – 2001)
0	127	126	1
3	139	139	0
4	148	148	0
5	159	158	1
6	170	169	1
7	181	183	–2

Table 6.3: Relationship of PSAE Reading Scores to ACT Reading Scale Scores

ACT Reading	PSAE Reading 2006	PSAE Reading 2001	Diff (2006 – 2001)
1	123	121	2
2	122	130	–8
3	128	128	0
4	121	120	1
5	126	127	–1
6	128	128	0
7	130	129	1
8	131	127	4
9	132	130	2
10	133	133	0
11	134	136	–2
12	137	139	–2
13	141	142	–1
14	144	146	–2
15	148	149	–1
16	150	150	0
17	152	153	–1
18	155	155	0
19	157	157	0
20	160	159	1
21	163	162	1
22	165	164	1
23	166	166	0
24	169	167	2
25	171	170	1
26	173	173	0
27	175	174	1
28	178	177	1
29	180	179	1
30	182	181	1
31	183	182	1
32	180	183	–3
33	185	184	1
34	187	186	1
35	190	188	2
36	192	190	2

Table 6.4: Relationship of PSAE Reading Scores to WorkKeys Reading for Information Level Scores

WK Reading	PSAE Reading 2006	PSAE Reading 2001	Diff (2006 – 2001)
0	124	125	–1
3	136	133	3
4	147	147	0
5	161	161	0
6	174	174	0
7	185	185	0

Table 6.5: Relationship of PSAE Science Scores to ACT Science Scale Scores

ACT Science	PSAE Science 2006	PSAE Science 2001	Diff (2006 – 2001)
1	125	120	5
2	NA	NA	NA
3	126	127	–1
4	125	NA	NA
5	124	123	1
6	127	125	2
7	129	127	2
8	129	127	2
9	131	129	2
10	133	130	3
11	134	132	2
12	137	134	3
13	139	136	3
14	141	139	2
15	138	142	–4
16	144	144	0
17	148	148	0
18	152	152	0
19	155	156	–1
20	158	160	–2
21	162	163	–1
22	165	166	–1
23	168	169	–1
24	172	173	–1
25	175	175	0
26	176	178	–2
27	179	180	–1
28	181	182	–1
29	183	184	–1
30	184	183	1
31	184	186	–2
32	186	184	2
33	186	188	–2
34	188	186	2
35	190	190	0
36	192	193	–1

Table 6.6: Relationship of PSAE Science Scores to ISBE Science Scale Scores

ISBE Science	PSAE Science 2006	PSAE Science 2001	Diff (2006 – 2001)
40	134	122	12
41	120	NA	NA
42	123	NA	NA
43	125	122	3
44	123	NA	NA
45	124	124	0
46	125	NA	NA
47	126	125	1
48	126	NA	NA
49	127	126	1
50	129	127	2
51	129	NA	NA
52	130	128	2
53	132	130	2
54	133	132	1
55	134	NA	NA
56	135	133	2
57	136	135	1
58	138	136	2
59	139	138	1
60	141	140	1
61	142	NA	NA
62	144	143	1
63	145	144	1
64	147	146	1
65	149	148	1
66	150	151	–1
67	152	153	–1
68	154	155	–1
69	156	157	–1
70	157	NA	NA
71	159	159	0
72	161	162	–1
73	163	164	–1
74	164	NA	NA
75	165	166	–1
76	167	168	–1
77	169	NA	NA
78	170	171	–1
79	171	173	–2
80	173	NA	NA
81	173	175	–2
82	175	NA	NA
83	176	177	–1
84	177	NA	NA
85	179	180	–1
86	179	NA	NA

Table 6.6: Relationship of PSAE Science Scores to ISBE Science Scale Scores

ISBE Science	PSAE Science 2006	PSAE Science 2001	Diff (2006 – 2001)
87	181	NA	NA
88	181	182	–1
89	183	NA	NA
90	183	NA	NA
91	185	185	0
92	186	NA	NA
93	186	NA	NA
94	189	NA	NA
95	188	187	1
96	187	NA	NA
97	192	NA	NA
98	191	NA	NA
99	NA	NA	NA
100	195	191	4

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 provide 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 is 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 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
Spring 2006 Results of the Prairie State Achievement Examination

This chapter provides a brief summary of the results of the spring 2006 PSAE administration. Individual and school PSAE reports from the spring 2006 administration were shipped to schools in May 2007. In addition to the PSAE reports, individual WorkKeys score reports for *Reading for Information* and *Applied Mathematics* were shipped to schools in August 2006 for distribution to students. Individual ACT reports had been mailed in May and June 2006 to students at their homes, along with ACT's standard student guide for interpreting scores. Students receive a Prairie State Achievement Award for any PSAE score or scores in the Exceeds Standards performance level.

PSAE Score Results

The results of the spring 2006 PSAE administration are based on the scores of the 131,690 grade 11 students who tested in April and May 2006. Table 8.1 shows the percentage of students in each of the four performance levels for the state for each of the three PSAE subjects. The percentage of students meeting or exceeding standards ranged from 51% to 58%, compared to 52% to 59% reported for spring 2005.

Table 8.2 on the following page shows the average score for the state for each of the three PSAE subjects, and the state average for the component assessments that make up each PSAE subject test. These values are the same as those reported for spring 2005.

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, and disability and migrant status. Results are provided only if five or more students are present in a given category. Again, the results are very similar to those reported for spring, 2005.

Tables 8.4 through 8.9 show the bivariate frequency between students' PSAE scores and their scores on the various PSAE components. For example, Table 8.5 provides the bivariate frequencies of students' PSAE reading scores and their ACT Reading scores. All PSAE scores range from 120 to 200, while the range of the component tests vary. As an example of how to read the tables, it can be seen in Table 8.5 that examinees with a PSAE Reading score of 158 had ACT Reading scores ranging from 11 (one examinee) to 31 (one examinee), with most examinees falling in the 15 to 22 range.

Table 8.1: Percentage of 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	8%	33%	44%	14%	58%
Mathematics	10%	37%	46%	8%	54%
Science	8%	41%	40%	11%	51%

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

Table 8.2: Average PSAE scores

PSAE Test	Score Range	Average Score
PSAE Reading	120–200	158
ACT Reading	1–36	20
WorkKeys *Reading for Information*	<3, 3–7	5
PSAE Mathematics	120–200	157
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

Table 8.3: Percentage of student scores within each PSAE performance level by various categories

	Mathematics				Reading				Science			
	Academic Warning	Below	Meets	Exceeds	Academic Warning	Below	Meets	Exceeds	Academic Warning	Below	Meets	Exceeds
All Students	10	37	46	8	8	33	44	14	8	41	40	11
Female	10	39	45	6	6	33	47	14	8	45	39	8
Male	10	34	46	10	11	34	42	14	9	37	41	14
Not Indicated	20	48	28	4	24	40	32	4	12	64	20	4
American Indian or Alaskan Native	11	40	46	4	10	37	42	12	8	46	39	8
Asian or Pacific Islander	3	20	54	23	4	23	48	25	3	28	48	21
Black, Non-Hispanic	25	54	20	1	16	51	31	3	21	62	16	1
Hispanic	15	52	32	1	13	48	34	4	14	59	26	2
White, Non-Hispanic	6	31	54	10	6	27	49	18	5	33	48	14
Multiracial/Ethnic	11	39	43	7	8	34	46	12	9	45	36	10
Not Indicated	13	39	41	7	11	36	40	13	11	44	37	8
Low Income	20	52	26	1	16	49	32	4	18	59	21	2
Not Low Income	6	31	53	10	6	27	49	18	5	34	47	14
LEP	21	43	33	3	25	47	24	4	25	48	22	5
Non-LEP	10	36	46	8	8	33	45	14	8	41	40	11
IEP	42	45	12	1	38	45	15	2	39	48	11	2
504	10	31	48	11	8	26	46	20	9	35	40	16
Migrant	21	64	14	0	21	50	29	0	14	79	0	7
Non-Migrant	10	36	46	8	8	33	45	14	8	41	40	11

Table 8.4: Bivariate frequency between PSAE Mathematics and ACT Mathematics

(ACT Mathematics scores 1–18)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
120	3	0	0	3	0	5	0	9	20	43	207	233	632	355	264	46	5	4
121	0	0	0	0	0	0	0	0	0	0	1	22	45	3	16	8	0	0
122	0	0	0	1	0	1	0	1	2	7	52	87	181	177	106	35	6	0
123	0	0	0	0	0	0	0	0	0	4	21	1	64	44	43	18	2	3
124	0	0	0	0	0	0	0	0	5	1	40	79	122	103	131	20	4	3
125	1	0	0	0	0	1	0	0	0	9	55	50	117	94	115	8	2	4
126	0	0	0	0	0	0	0	0	3	0	18	46	67	83	0	1	6	0
127	0	0	0	0	0	0	0	2	0	4	40	31	196	147	171	37	8	3
128	0	0	0	0	0	2	0	0	2	3	32	103	186	87	149	25	5	1
129	0	0	0	1	0	0	0	0	0	5	18	48	232	297	93	44	3	5
130	0	0	0	0	0	0	0	2	0	0	6	6	152	79	289	34	5	5
131	0	0	0	0	0	0	0	0	0	3	24	132	73	268	172	51	6	3
132	0	0	0	0	0	0	0	3	0	14	65	145	480	292	521	70	11	4
133	0	0	0	0	0	0	0	0	0	0	1	3	8	304	212	75	8	0
134	0	0	0	0	0	0	0	2	0	0	25	66	490	348	424	109	11	3
135	0	0	0	0	0	0	0	0	0	3	31	60	116	321	658	159	27	1
136	0	0	0	0	0	0	0	1	1	0	6	62	290	388	342	213	25	7
137	0	0	0	0	0	0	0	0	0	2	19	58	103	307	671	233	20	5
138	0	0	0	0	0	0	0	0	0	0	12	40	159	390	716	177	74	9
139	0	0	0	0	0	0	0	1	0	3	9	39	230	306	733	434	16	4
140	0	0	0	0	0	0	0	0	0	0	2	23	127	341	887	244	77	19
141	0	0	0	0	0	0	0	0	0	0	10	36	212	90	787	367	54	6
142	0	0	0	0	0	0	0	0	0	0	1	20	205	543	857	958	243	36
143	0	0	0	0	0	0	0	0	0	0	3	24	23	235	1281	635	82	25
144	0	0	0	0	0	0	0	0	0	0	2	2	144	220	1160	680	306	61
145	0	0	0	0	0	0	0	0	0	0	3	19	29	144	439	1179	321	68
146	0	0	0	0	0	0	0	0	0	0	0	10	71	206	1001	791	249	167
147	0	0	0	0	0	0	0	0	0	0	1	0	23	171	741	1412	713	179
148	0	0	0	0	0	0	0	0	0	0	0	6	18	67	689	797	292	121
149	0	0	0	0	0	0	0	0	0	0	0	0	30	71	628	896	679	453
150	0	0	0	0	0	0	0	0	0	0	0	0	10	72	306	1051	753	377
151	0	0	0	0	0	0	0	0	0	0	0	1	8	14	277	1050	830	271
152	0	0	0	0	0	0	0	0	0	0	0	0	8	25	178	625	861	562
153	0	0	0	0	0	0	0	0	0	0	0	0	3	18	213	732	804	885
154	0	0	0	0	0	0	0	0	0	0	0	0	2	2	178	620	422	365
155	0	0	0	0	0	0	0	0	0	0	0	0	3	7	76	335	1077	714
156	0	0	0	0	0	0	0	0	0	0	0	0	0	5	53	338	641	821
157	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	330	610	835
158	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	152	341	754
159	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	84	457	1030
160	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	24	255	473
161	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	158	512
162	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	61	489
163	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	241

Table 8.4: Bivariate frequency between PSAE Mathematics and ACT Mathematics

(ACT Mathematics scores 1–18)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
164	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	139
165	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27
166	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
167	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
168	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
169	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
171	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
172	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
173	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
174	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
176	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
177	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
178	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
179	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
180	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
181	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
182	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
183	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
184	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
185	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
187	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
188	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
189	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
190	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
191	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
192	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
193	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
194	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
195	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
196	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
197	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
198	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mn*	121	NA	NA	122	NA	123	NA	126	122	125	127	130	132	136	141	147	152	155

Table 8.4: Bivariate frequency between PSAE Mathematics and ACT Mathematics

(ACT Mathematics scores 19–36)

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
124	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
125	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
126	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
127	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
128	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
129	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
130	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
131	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
132	10	2	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0
133	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
134	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0
135	7	0	1	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0
136	4	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
137	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
138	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
139	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
140	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
141	2	0	3	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
142	7	3	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
143	6	0	1	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0
144	10	0	1	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0
145	11	4	4	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0
146	20	7	0	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0
147	18	13	4	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
148	64	11	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
149	85	21	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
150	53	16	11	9	0	1	0	0	0	1	0	0	0	0	0	0	0	0
151	177	74	9	6	4	0	0	1	0	0	0	0	0	0	0	0	0	0
152	183	60	38	10	6	0	1	0	1	0	0	0	0	0	0	0	0	0
153	148	40	49	30	3	4	0	0	1	0	0	0	0	0	0	0	0	0
154	266	196	20	32	12	1	0	0	0	0	0	0	0	0	0	0	0	0
155	492	183	141	22	11	6	1	0	0	1	0	0	0	0	0	0	0	0
156	308	153	102	78	8	9	3	1	1	0	0	0	0	0	0	0	0	0
157	728	478	111	76	58	10	9	0	0	0	1	0	0	0	0	0	0	0
158	453	321	293	61	50	44	6	1	2	2	0	0	1	0	0	0	0	0
159	416	278	263	250	46	45	17	8	0	0	0	0	0	1	0	0	0	0
160	823	726	238	193	121	52	24	4	5	1	0	0	0	0	0	0	0	0
161	332	400	617	181	133	121	17	1	8	0	0	1	0	0	0	0	0	0
162	607	803	418	540	129	110	49	11	7	0	1	0	0	0	0	0	0	0
163	463	410	397	339	380	145	65	11	5	2	0	0	0	0	0	0	0	0

Table 8.4: Bivariate frequency between PSAE Mathematics and ACT Mathematics

(ACT Mathematics scores 19–36)

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
164	296	678	792	305	302	306	68	11	16	2	0	0	1	0	0	0	0	0
165	166	491	755	748	266	331	151	32	4	3	2	0	0	0	0	0	0	0
166	86	380	583	727	720	362	65	37	32	4	0	2	0	0	0	0	0	0
167	25	119	487	699	727	820	356	41	71	5	3	1	0	0	0	0	0	0
168	1	56	175	460	688	690	392	153	15	3	4	1	0	0	0	0	0	0
169	0	4	72	368	565	1033	711	270	89	24	1	4	1	0	0	0	0	0
170	0	0	25	112	475	990	531	224	122	29	14	1	0	0	0	0	0	0
171	0	0	1	34	148	781	905	177	133	27	12	5	0	0	0	0	0	1
172	0	0	1	5	65	535	917	545	317	101	11	4	1	2	0	0	0	0
173	0	0	0	3	10	288	799	592	535	99	31	12	1	0	1	0	0	0
174	0	0	0	0	0	101	489	580	466	88	26	7	0	2	0	1	0	0
175	0	0	0	0	0	2	257	497	835	305	112	35	2	1	1	0	0	1
176	0	0	0	0	0	1	32	131	878	213	118	33	0	1	0	1	1	0
177	0	0	0	0	0	0	2	228	705	458	269	34	6	0	5	0	0	0
178	0	0	0	0	0	0	0	1	214	540	141	80	27	4	0	1	0	0
179	0	0	0	0	0	0	0	1	314	520	756	193	0	19	7	0	3	0
180	0	0	0	0	0	0	0	0	2	288	288	366	55	0	0	4	0	0
181	0	0	0	0	0	0	0	0	1	87	216	206	102	41	10	0	5	0
182	0	0	0	0	0	0	0	0	0	3	285	283	1	0	0	1	1	0
183	0	0	0	0	0	0	0	0	0	2	98	531	154	78	26	12	7	0
184	0	0	0	0	0	0	0	0	0	0	3	362	265	167	66	0	5	0
185	0	0	0	0	0	0	0	0	0	0	1	116	1	3	1	1	8	1
186	0	0	0	0	0	0	0	0	0	0	0	4	241	251	131	40	22	2
187	0	0	0	0	0	0	0	0	0	0	0	0	124	3	0	0	0	0
188	0	0	0	0	0	0	0	0	0	0	0	4	0	255	244	107	10	0
189	0	0	0	0	0	0	0	0	0	0	0	0	3	175	0	1	71	5
190	0	0	0	0	0	0	0	0	0	0	0	0	0	0	242	179	37	0
191	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	114	17
192	0	0	0	0	0	0	0	0	0	0	0	0	2	0	155	1	0	0
193	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	241	110	0
194	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6	0	1	40
195	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	162	213	0
196	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	324	0
197	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3	86
198	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	160	0
199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	138
Mn*	158	161	163	164	166	168	170	172	175	177	179	181	184	186	188	191	194	197

Note: N=116,960(any examinee not having PSAE scores was deleted). * is rounded conditional mean given ACT Mathematics scores. NA: Not applicable

Table 8.5: Bivariate frequency between PSAE Reading and ACT Reading

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
120	3	4	2	2	12	31	53	89	162	201	430	485	374	136	88	51	34	33	22	9	11	1	2	4	0	1	1	0	0	0	0	0	0	0	0	0
121	0	1	0	0	0	2	0	0	29	0	25	164	23	2	29	2	4	2	3	3	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0
122	0	0	0	1	0	0	19	0	0	75	44	45	40	20	9	7	1	7	1	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	0	0	2	3	1	30	10	0	75	108	125	0	14	9	10	0	1	2	1	0	1	4	1	0	0	0	0	0	0	0	0	0	0	0
124	0	0	2	0	0	8	3	0	56	20	57	75	73	25	0	9	7	3	0	1	0	1	0	2	0	1	0	0	0	0	0	0	0	0	0	0
125	0	0	0	0	1	1	14	1	28	59	42	141	55	31	35	0	9	3	0	1	4	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0
126	0	0	0	0	0	11	10	45	0	29	148	69	143	35	14	12	0	5	4	1	0	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0
127	0	0	1	0	0	2	14	2	63	5	59	185	97	4	27	13	11	0	11	3	1	1	0	0	0	1	0	2	0	0	0	0	0	1	0	0
128	0	0	0	0	2	0	0	43	2	78	181	186	6	83	0	22	9	6	0	0	1	1	4	0	2	0	0	1	0	0	0	0	0	0	0	0
129	0	0	2	0	0	12	11	1	71	38	0	98	173	0	105	0	14	5	0	0	3	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0
130	0	1	0	0	1	0	13	1	14	120	48	357	161	69	1	43	0	8	8	1	1	3	0	1	1	1	0	1	0	0	0	0	0	0	0	0
131	0	0	0	0	0	8	5	58	2	29	213	2	112	72	43	0	27	0	9	0	1	4	4	1	0	0	1	0	0	0	0	0	0	0	0	0
132	0	0	1	0	0	0	18	1	96	5	55	303	368	3	100	29	0	8	6	3	0	0	2	0	0	1	0	1	0	0	0	0	0	0	0	0
133	1	0	0	0	5	1	0	53	3	116	78	308	98	205	3	30	0	20	4	0	5	0	2	1	0	0	0	0	1	0	0	0	0	0	0	0
134	0	0	1	0	0	6	27	0	99	22	219	92	324	10	152	0	34	0	7	1	2	1	0	0	0	0	2	1	1	0	0	0	0	1	0	0
135	0	0	0	0	2	2	0	0	13	130	56	357	225	96	76	36	1	22	0	5	3	0	0	1	2	1	0	0	0	0	0	0	0	0	0	0
136	0	0	1	0	0	7	8	50	9	19	255	312	114	201	63	46	27	19	7	0	0	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0
137	0	0	0	0	1	0	22	0	103	5	40	371	409	11	168	51	53	23	22	0	1	0	1	1	0	1	1	1	2	0	0	0	0	0	0	0
138	0	0	0	0	0	4	4	35	2	148	50	380	331	94	146	1	2	40	14	11	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
139	0	0	0	0	1	1	18	0	70	13	221	78	557	288	84	117	31	34	13	0	14	4	2	4	0	0	0	0	0	0	0	0	0	0	0	0
140	0	0	0	0	0	2	2	30	5	130	32	729	392	98	458	74	104	79	25	8	4	3	2	0	1	0	0	0	1	0	0	0	0	0	0	0
141	0	0	0	0	0	2	12	1	59	13	244	45	520	426	65	191	44	54	13	18	5	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0
142	0	0	0	0	0	3	0	18	1	119	21	327	567	7	383	61	139	35	51	0	14	0	4	0	4	0	0	0	0	0	0	0	0	0	0	0
143	0	0	0	0	0	0	4	1	52	8	188	411	89	552	342	290	47	137	31	35	7	4	0	1	1	0	0	0	0	0	1	0	0	0	0	0
144	0	0	0	0	0	0	0	13	0	91	14	260	528	15	509	4	1	110	23	1	4	3	1	0	0	2	0	1	0	0	0	0	0	0	0	0
145	0	0	0	0	0	1	3	0	27	6	155	381	564	75	476	71	249	246	78	18	19	7	7	1	0	0	0	0	0	1	0	0	0	0	0	0
146	0	0	0	0	0	1	0	5	1	55	6	246	468	550	637	428	42	183	36	56	29	16	9	2	2	1	0	0	0	0	0	0	0	0	0	0
147	0	0	0	0	0	0	2	0	15	0	85	283	557	72	566	70	348	44	161	1	46	5	3	6	0	2	0	0	0	0	0	1	0	0	0	0
148	0	0	0	0	0	0	0	0	0	26	6	166	366	570	59	520	60	336	44	129	41	0	10	2	2	0	1	0	0	0	0	0	0	0	0	0
149	0	0	0	0	0	0	0	2	0	0	56	180	409	53	1237	42	515	320	36	0	100	26	3	9	3	3	0	1	0	1	0	0	0	0	0	0
150	0	0	0	0	0	0	0	0	9	1	4	114	317	500	43	625	61	481	273	31	0	15	37	7	0	0	0	0	0	0	0	0	0	0	0	0
151	0	0	0	0	0	0	0	1	0	20	36	121	363	37	567	32	575	385	39	183	104	1	15	21	2	1	1	0	0	0	0	0	0	0	0	0
152	0	0	0	0	0	0	0	0	3	5	1	51	175	403	664	609	29	586	433	34	177	66	6	9	1	5	0	2	0	0	0	0	0	0	0	0
153	0	0	0	0	0	0	0	0	0	0	14	61	222	31	497	30	630	517	5	324	163	1	41	7	12	2	2	0	0	0	0	0	0	0	0	0
154	0	0	0	0	0	0	0	0	2	1	2	23	97	288	565	14	27	646	515	0	295	111	20	31	3	0	0	5	1	0	0	0	0	0	0	0
155	0	0	0	0	0	1	0	1	0	2	0	36	118	14	365	568	11	610	41	435	203	29	81	28	0	12	1	4	0	0	0	0	0	0	0	0
156	0	0	0	0	0	0	0	0	0	0	7	11	57	149	413	12	694	649	625	0	397	185	9	75	16	1	5	0	1	0	0	0	0	0	0	0
157	0	0	0	0	0	0	0	1	0	1	0	3	94	12	243	517	11	672	35	581	319	18	158	52	3	15	0	4	0	0	0	0	0	0	0	0
158	0	0	0	0	0	0	0	0	0	0	1	14	0	101	296	0	592	7	32	2	488	291	24	123	44	4	11	3	1	2	0	1	0	0	0	0
159	0	0	0	0	0	0	0	0	0	0	0	1	54	2	140	381	0	655	669	608	438	23	235	75	6	33	0	3	2	0	0	0	0	0	0	0
160	0	0	0	0	0	0	0	0	1	0	0	5	1	0	185	0	421	647	31	4	31	3	24	9	2	7	5	3	0	0	0	0	0	0	1	0
161	0	0	0	0	0	0	0	0	0	0	0	3	7	50	71	253	0	488	632	23	573	373	22	184	78	3	25	2	5	0	4	0	0	1	0	0
162	0	0	0	0	0	0	0	0	0	0	0	1	14	0	94	0	289	541	14	674	553	10	356	170	3	61	8	12	0	0	0	0	0	0	0	0
163	0	0	0	0	0	0	0	0	0	0	0	0	2	16	35	160	0	332	539	0	672	507	23	300	125	11	29	13	1	2	0	0	0	0	0	0

Table 8.5: Bivariate frequency between PSAE Reading and ACT Reading

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
164	0	0	0	0	0	0	0	0	0	0	0	0	7	0	37	0	162	412	6	618	560	8	492	240	4	95	7	29	10	2	2	0	0	0	0	0
165	0	0	0	0	0	0	0	0	0	0	0	0	0	6	11	71	0	225	493	0	647	565	15	414	212	13	67	23	2	5	0	0	0	0	0	0
166	0	0	0	0	0	0	0	0	0	0	0	0	0	3	22	0	80	252	4	515	606	21	510	355	5	186	9	52	13	3	2	0	1	1	0	0
167	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	24	0	112	298	0	515	616	19	505	339	12	134	30	4	14	1	0	2	0	0	0
168	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	36	132	1	320	530	12	611	442	5	257	10	108	32	3	8	1	0	0	0	0
169	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	33	187	0	366	565	10	592	429	14	231	73	1	15	4	0	7	0	0	0
170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	55	1	179	439	2	575	593	1	395	5	189	62	2	16	2	0	3	0	0
171	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	9	76	0	243	446	4	554	493	15	320	7	4	0	1	0	2	0	0	0
172	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	15	0	88	238	2	472	567	4	487	12	454	101	39	1	2	10	0	0	0
173	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	19	0	114	271	5	529	578	11	459	10	8	2	36	1	3	4	0	1
174	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	35	119	1	307	532	7	550	7	586	4	87	1	5	23	0	2	0
175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	33	154	1	351	499	8	0	14	205	2	69	0	1	8	0	0
176	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	55	0	166	371	11	525	550	355	5	164	1	4	37	1	3	2
177	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	56	0	189	358	9	0	467	279	0	108	0	5	20	0	0
178	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	70	239	3	3	521	431	0	222	5	7	82	2	7	1
179	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	15	1	83	230	394	0	478	400	0	202	1	7	37	0	0
180	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	104	1	2	415	431	0	334	0	11	134	0	0	7
181	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	22	122	289	3	447	442	0	305	6	11	91	20	0
182	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	25	0	3	255	429	0	440	0	5	235	3	1	5
183	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	24	110	3	301	425	1	362	0	2	161	53	3
184	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	27	158	328	3	428	2	6	324	3	91	21
185	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	40	147	334	1	402	0	8	259	1	0
186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	210	2	356	1	7	359	6	1	2
187	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	158	1	326	0	10	2	161	31
188	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	1	205	2	3	337	285	1	5
189	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	65	1	229	1	2	339	213	62
190	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	11	88	2	1	258	0	0	4
191	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92	2	0	8	272	112
192	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	1	100	227	1	2
193	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	130	218	158
194	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	109	5
195	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	143
196	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	96
197	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
198	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18
Mn*	123	122	128	121	126	128	130	131	132	133	134	137	141	144	148	150	152	155	157	160	163	165	166	169	171	173	175	178	180	182	183	180	185	187	190	192

Note: N=116,960(any examinee not having PSAE scores was deleted). * is rounded conditional mean given ACT Reading scores.

Table 8.6: Bivariate frequency between PSAE Science and ACT Science

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
120	3	0	1	0	2	6	7	16	20	17	16	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121	0	0	0	0	0	2	0	7	5	19	14	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122	0	0	0	0	0	3	4	8	11	29	27	10	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	0	1	1	1	10	19	31	42	66	23	9	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
124	0	0	0	1	0	2	8	23	24	38	75	30	15	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
125	0	0	0	0	0	1	5	26	33	64	109	60	32	20	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
126	0	0	1	1	1	2	7	8	47	88	164	88	59	25	12	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
127	0	0	0	1	1	0	7	21	21	125	74	6	53	40	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
128	0	0	0	0	0	7	2	22	50	93	152	150	30	18	24	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
129	0	0	0	0	1	2	8	27	52	179	219	159	121	79	28	19	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
130	0	0	0	0	0	2	7	9	57	89	272	217	103	76	40	36	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
131	0	0	1	0	0	0	5	17	3	207	134	89	193	117	57	40	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
132	0	0	0	0	0	1	7	28	71	232	244	243	250	179	55	87	27	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
133	0	0	0	0	0	3	0	21	30	143	281	283	294	83	78	68	31	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
134	0	0	0	0	0	0	8	15	44	75	309	316	85	242	75	170	71	13	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
135	0	0	0	0	0	0	2	2	38	202	51	62	330	295	2	213	35	10	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
136	0	0	0	0	0	0	9	10	35	182	250	264	347	340	70	100	121	28	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
137	0	0	0	0	0	0	1	7	6	165	180	257	338	18	76	271	163	35	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
138	0	0	0	0	0	1	0	6	20	94	207	334	69	375	69	328	31	36	9	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
139	0	0	0	0	0	1	2	4	18	128	183	290	385	425	58	131	296	86	15	9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
140	0	0	0	0	0	0	0	3	10	45	163	33	377	424	42	330	381	118	7	10	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
141	1	0	0	0	0	1	6	4	16	88	24	231	382	51	60	421	527	183	36	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
142	0	0	0	0	0	0	2	5	10	66	112	233	317	436	46	431	40	36	37	16	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
143	0	0	0	0	0	0	0	0	7	70	81	175	289	427	49	457	668	279	22	33	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
144	0	0	0	0	0	0	0	1	6	42	74	145	25	363	37	87	728	353	70	28	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
145	0	0	0	0	0	0	0	1	3	27	45	112	246	367	5	492	777	466	92	42	8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
146	0	0	0	0	0	0	0	1	1	17	46	98	200	322	36	487	835	257	134	88	27	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0
147	0	0	0	0	0	0	0	1	5	13	34	49	168	235	31	441	466	331	29	146	11	6	3	0	0	0	0	0	0	0	0	0	0	0	0	0
148	0	0	0	0	0	0	0	0	5	5	29	5	124	28	27	433	454	703	211	76	33	8	5	0	0	1	0	0	0	0	0	0	0	0	0	0
149	0	0	0	0	0	0	0	0	0	2	15	42	99	224	18	306	979	802	215	145	57	7	0	2	1	0	0	0	0	0	0	0	0	0	0	0
150	0	0	0	0	0	0	0	1	1	2	11	26	58	164	18	323	855	864	306	297	87	15	3	1	1	0	0	0	0	0	0	0	0	0	0	0
151	0	0	0	0	0	0	0	0	0	2	5	17	53	133	15	31	788	979	344	340	54	24	4	1	0	0	0	0	0	0	0	0	0	0	0	0
152	0	0	0	0	0	0	0	0	0	1	2	13	30	90	11	214	722	50	412	489	114	36	9	4	3	0	2	0	0	0	0	0	0	0	0	0
153	0	0	0	0	0	0	0	0	0	2	1	11	18	78	3	189	307	994	55	523	189	23	9	1	0	1	0	0	0	0	0	0	0	0	0	0
154	0	0	0	0	0	0	0	0	0	2	0	6	17	53	8	142	562	905	488	280	281	53	21	4	0	0	0	0	0	0	0	0	0	0	0	0
155	0	0	0	0	0	0	0	0	0	0	2	2	8	27	7	99	477	858	499	374	333	106	9	0	3	0	0	0	0	0	0	0	0	0	0	0
156	0	0	0	0	0	0	0	0	0	1	1	2	8	21	5	77	395	825	516	798	200	123	23	8	2	0	0	0	0	0	0	0	0	0	0	0
157	0	0	0	0	0	0	0	0	0	0	0	1	0	10	4	38	224	720	505	799	547	221	54	10	1	1	0	0	0	0	0	0	0	0	0	0
158	0	0	0	0	0	0	0	0	0	0	0	1	1	3	2	36	207	648	394	855	309	298	77	2	3	0	0	2	0	1	0	0	0	0	0	0
159	0	0	0	0	0	0	0	0	0	0	0	0	2	4	0	17	168	515	423	869	653	160	144	19	8	1	0	0	0	0	0	0	0	0	0	0
160	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	13	110	410	371	843	749	231	163	18	13	0	0	1	0	0	1	0	0	0	0	0
161	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	7	78	299	301	805	863	470	74	47	15	0	0	0	0	0	0	0	0	0	0	0
162	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	47	219	24	704	885	568	247	78	8	3	1	0	0	0	0	0	0	0	0	0
163	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	120	247	386	861	602	186	104	28	8	0	0	0	0	0	0	0	0	0	0

Table 8.6: Bivariate frequency between PSAE Science and ACT Science

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
164	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	17	35	158	288	784	724	411	53	24	7	4	0	0	0	0	0	0	0	0	0
165	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	82	132	519	667	811	469	205	60	8	2	2	0	0	0	0	0	0	0	0
166	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	31	92	377	646	790	609	311	87	1	7	1	0	0	0	0	0	0	0	0
167	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	17	69	286	541	821	681	203	124	18	4	0	0	0	0	0	0	0	0	0
168	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	31	209	267	674	735	388	210	32	18	6	0	0	0	0	0	0	0	0
169	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	13	143	458	629	685	473	268	47	19	6	3	0	1	0	0	0	0	0
170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7	74	206	553	701	600	139	82	43	0	2	0	1	1	0	0	0	0
171	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	49	6	379	659	602	335	90	55	4	0	0	1	0	0	0	0	0
172	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	14	161	301	601	668	428	112	88	9	5	0	0	3	0	0	0	0
173	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	15	71	495	642	289	157	169	16	5	0	1	1	0	0	0	0
174	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	58	233	379	676	531	201	211	29	12	0	3	1	0	0	0	0
175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	10	82	275	517	564	238	236	58	23	1	5	1	0	2	0	0
176	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	33	189	487	620	274	306	73	31	0	14	3	1	0	1	0
177	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	116	368	541	285	380	105	51	1	13	7	0	6	0	0
178	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	57	263	513	277	456	112	78	1	27	10	0	0	0	0
179	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	25	150	444	10	488	188	76	1	40	20	0	4	2	0
180	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	92	320	415	489	204	127	1	78	27	1	19	3	0
181	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	42	217	150	432	208	145	1	86	53	0	8	5	2
182	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	121	107	377	202	164	2	96	63	3	24	4	1
183	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	64	59	295	211	202	3	132	76	0	38	19	1
184	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	31	30	184	171	197	3	157	104	0	38	21	2
185	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	8	88	127	161	6	186	117	1	59	37	6
186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	67	108	129	0	162	145	1	77	49	10
187	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	21	41	102	2	123	147	3	125	49	10
188	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	67	3	100	129	0	117	3	13
189	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	3	0	0	117	2	114	61	31
190	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	32	1	50	0	0	0	63	0
191	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	27	81	3	102	1	33
192	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0
193	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	42	0	82	75	42
194	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	58	0
195	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	35
196	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	39	1	0
197	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	34	0
198	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	24
199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
Mn*	125	NA	126	125	124	127	129	129	131	133	134	137	139	141	138	144	148	152	155	158	162	165	168	172	175	176	179	181	183	184	184	186	186	188	190	192

Note: N=116,960 (any examinee not having PSAE scores was deleted). * is rounded conditional mean given ACT Science scores. NA: Not applicable

Table 8.7: Bivariate frequency between PSAE Mathematics and WorkKeys *Applied Mathematics*

	0	3	4	5	6	7
120	1828	1	0	0	0	0
121	95	0	0	0	0	0
122	656	3	0	0	0	0
123	200	0	0	0	0	0
124	504	5	0	0	0	0
125	447	11	0	0	0	0
126	204	21	0	0	0	0
127	597	46	0	0	0	0
128	522	73	0	0	0	0
129	615	133	0	0	0	0
130	563	19	0	0	0	0
131	502	232	0	0	0	0
132	831	793	0	0	0	0
133	298	316	0	0	0	0
134	490	990	2	0	0	0
135	198	1186	3	0	0	0
136	216	1118	8	0	0	0
137	70	1330	21	0	0	0
138	71	1454	54	0	0	0
139	19	1680	79	0	0	0
140	22	1542	159	0	0	0
141	13	1288	268	0	0	0
142	22	2072	783	0	0	0
143	10	1249	1057	3	0	0
144	4	1048	1533	4	0	0
145	12	706	1468	39	0	0
146	4	441	2017	64	0	0
147	4	379	2725	172	0	0
148	0	195	1468	404	0	0
149	0	196	1978	696	0	0
150	1	82	1557	1020	0	0
151	1	93	1283	1345	0	0
152	1	55	866	1636	0	0
153	0	42	803	2080	5	0
154	0	14	548	1550	4	0
155	0	8	577	2466	18	0
156	0	7	361	2097	56	0
157	0	1	278	2905	83	0
158	0	6	166	2156	158	0
159	0	1	120	2478	298	1
160	0	1	90	2398	450	1
161	0	1	31	2070	386	1
162	0	0	25	2378	819	3
163	0	0	8	1738	724	8
164	0	0	3	1796	1107	12
165	0	0	3	1535	1412	26
166	0	0	2	1217	1761	21
167	0	0	0	1296	2010	48
168	0	0	0	572	1966	100
169	0	0	0	549	2450	143
170	0	0	0	168	2106	249
171	0	0	1	173	1868	182
172	0	0	0	117	2046	341
173	0	0	0	46	1847	478
174	0	0	0	35	1138	587
175	0	0	0	40	1254	754
176	0	0	0	3	967	439
177	0	0	0	11	763	933
178	0	0	0	4	518	486
179	0	0	0	11	696	1106
180	0	0	0	4	424	575
181	0	0	0	5	356	307
182	0	0	0	0	5	569
183	0	0	0	0	275	633
184	0	0	0	0	238	630
185	0	0	0	0	11	121
186	0	0	0	0	196	495
187	0	0	0	0	0	127
188	0	0	0	0	117	503
189	0	0	0	0	76	179
190	0	0	0	0	37	421
191	0	0	0	0	17	115
192	0	0	0	0	0	158
193	0	0	0	0	0	351
194	0	0	0	0	0	49
195	0	0	0	0	0	377
196	0	0	0	0	0	328
197	0	0	0	0	0	92
198	0	0	0	0	0	167
199	0	0	0	0	0	7
200	0	0	0	0	0	141
Mn*	127	139	148	159	170	181

Note: N=116,960 (any examinee not having PSAE scores was deleted). * is rounded conditional mean given WorkKeys mathematics scores.

Table 8.8: Bivariate frequency between PSAE Reading and WorkKeys *Reading for Information*

	0	3	4	5	6	7
120	2239	2	0	0	0	0
121	291	1	0	0	0	0
122	271	1	0	0	0	0
123	391	6	0	0	0	0
124	330	13	0	0	0	0
125	384	45	0	0	0	0
126	435	95	0	0	0	0
127	357	146	0	0	0	0
128	319	308	0	0	0	0
129	304	231	2	0	0	0
130	296	556	2	0	0	0
131	162	414	15	0	0	0
132	150	819	31	0	0	0
133	63	807	64	0	0	0
134	50	787	165	0	0	0
135	34	784	210	0	0	0
136	31	756	356	0	0	0
137	30	709	548	0	0	0
138	14	613	638	0	0	0
139	24	560	965	1	0	0
140	11	742	1424	2	0	0
141	5	325	1373	14	0	0
142	8	239	1485	22	0	0
143	3	214	1927	57	0	0
144	3	140	1332	105	0	0
145	2	129	2079	175	0	0
146	3	111	2392	267	0	0
147	3	60	1811	393	0	0
148	1	55	1757	525	0	0
149	2	44	2289	661	0	0
150	0	44	1521	952	1	0
151	0	25	1335	1141	3	0
152	0	17	1331	1902	9	0
153	0	16	1058	1467	18	0
154	0	6	1011	1601	28	0
155	0	11	820	1685	43	1
156	0	6	719	2506	75	0
157	0	4	587	2041	106	1
158	0	7	519	1391	120	0
159	0	2	373	2747	203	0
160	0	1	50	1130	198	1
161	0	5	318	2079	392	3
162	0	0	250	2144	404	2
163	0	0	180	2037	548	2
164	0	0	145	1916	623	7
165	0	0	110	1848	794	17
166	0	0	81	1672	862	25
167	0	0	51	1625	925	26
168	0	0	61	1405	1006	43
169	0	0	27	1339	1128	36
170	0	0	23	1235	1205	67
171	0	0	3	834	1252	87
172	0	0	12	1085	1292	106
173	0	0	8	516	1394	136
174	0	0	3	708	1400	159
175	0	0	0	286	872	193
176	0	0	2	567	1460	226
177	0	0	0	409	837	254
178	0	0	0	314	969	325
179	0	0	0	241	1278	331
180	0	0	0	141	1191	122
181	0	0	0	111	1212	436
182	0	0	0	6	1111	287
183	0	0	0	56	951	439
184	0	0	0	20	855	518
185	0	0	0	0	670	529
186	0	0	0	0	375	571
187	0	0	0	0	204	491
188	0	0	0	0	297	589
189	0	0	0	0	280	636
190	0	0	0	0	4	372
191	0	0	0	0	112	374
192	0	0	0	0	1	344
193	0	0	0	0	0	531
194	0	0	0	0	0	138
195	0	0	0	0	0	166
196	0	0	0	0	0	119
197	0	0	0	0	0	0
198	0	0	0	0	0	0
199	0	0	0	0	0	0
200	0	0	0	0	0	18
Mn*	124	136	147	161	174	185

Note: N=116,960 (any examinee not having PSAE scores was deleted). * is rounded conditional mean given WorkKeys Reading for Information scores.

Table 8.9: Bivariate frequency between PSAE Science and ISBE-developed Science

(ISBE-developed science scores 40–57)

	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
120	0	4	4	5	19	6	18	13	6	11	5	3	5	0	0	0	0	0
121	0	0	0	0	2	4	7	10	14	6	2	4	1	0	0	0	0	0
122	0	0	0	2	3	6	5	15	13	25	9	11	2	4	0	0	0	0
123	0	0	1	0	5	10	16	20	17	33	43	31	17	11	2	1	1	0
124	0	0	0	0	1	7	7	20	27	34	30	27	29	19	12	5	1	0
125	0	0	0	1	4	7	12	27	22	46	42	65	67	29	10	21	2	1
126	0	0	0	1	0	3	9	30	14	48	55	73	81	100	51	22	9	6
127	0	0	0	1	1	0	0	2	9	30	31	22	20	60	80	53	21	14
128	0	0	0	0	4	1	5	22	14	13	17	70	102	67	65	82	41	31
129	0	0	0	0	0	2	4	11	10	31	48	78	94	110	106	123	112	91
130	0	0	0	0	2	0	1	6	9	25	26	53	56	99	89	144	151	116
131	0	0	0	0	0	4	0	5	7	10	17	51	80	97	99	86	84	99
132	0	0	1	0	1	0	2	12	9	20	26	38	44	97	116	203	176	166
133	0	0	0	0	0	0	1	5	2	18	30	17	31	16	60	83	177	216
134	1	0	0	0	0	0	0	1	3	11	19	40	56	81	73	117	90	100
135	0	0	0	1	0	0	0	2	1	6	5	17	23	39	74	122	142	146
136	0	0	0	0	0	0	1	0	2	4	17	19	24	51	65	50	76	134
137	0	0	0	0	0	0	2	0	2	2	7	15	28	31	46	111	140	110
138	0	0	0	1	0	0	1	0	0	3	11	3	16	21	16	13	21	122
139	0	0	0	0	0	0	1	0	0	3	2	15	17	42	58	98	125	91
140	0	0	0	0	0	0	0	0	1	0	3	2	13	25	24	69	117	139
141	0	0	0	0	0	0	0	0	0	0	0	3	10	16	28	58	82	120
142	0	0	0	0	0	0	0	3	0	0	1	4	4	14	12	17	11	19
143	0	0	0	0	0	0	0	0	0	0	0	1	2	7	8	35	65	112
144	0	0	0	0	0	0	0	0	0	0	0	1	3	6	12	25	50	64
145	0	0	0	0	0	0	0	0	0	0	0	0	1	2	7	8	35	39
146	0	0	0	0	0	0	0	0	0	0	0	1	1	2	6	16	29	32
147	0	0	0	0	0	0	0	0	0	0	0	0	1	1	3	6	10	25
148	0	0	0	0	0	0	0	0	0	1	0	1	0	3	1	2	5	20
149	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	1	7	15
150	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	1	9
151	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	2
152	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	3	2	5
153	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
154	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
155	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
156	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
157	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
158	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0
159	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
160	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
161	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
162	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
163	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 8.9: Bivariate frequency between PSAE Science and ISBE-developed Science

(ISBE-developed science scores 40–57)

	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
164	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
165	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
166	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
167	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
168	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
169	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
171	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
172	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
173	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
174	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
176	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
177	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
178	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
179	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
180	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
181	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
182	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
183	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
184	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
185	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
187	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
188	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
189	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
190	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
191	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
192	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
193	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
194	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
195	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
196	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
197	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
198	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mn*	134	120	123	125	123	124	125	126	126	127	129	129	130	132	133	134	135	136

Note: N=116,960 (any examinee not having PSAE scores was deleted). *is rounded conditional mean given ISBE science scores.

Table 8.9: Bivariate frequency between PSAE Science and ISBE-developed Science

(ISBE-developed science scores 58–74)

	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
125	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
126	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
127	7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
128	22	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
129	41	25	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0
130	73	39	20	3	3	0	0	0	0	0	0	0	0	0	0	0	0
131	131	65	20	7	4	1	0	0	0	0	0	0	0	0	0	0	0
132	163	120	116	92	19	7	0	0	0	0	0	0	0	0	0	0	0
133	219	182	131	90	26	14	2	0	0	0	0	0	0	0	0	0	0
134	118	236	193	162	70	34	16	3	0	0	0	0	0	0	0	0	0
135	239	88	52	70	110	70	25	13	0	0	0	0	0	0	0	0	0
136	217	245	222	203	186	124	65	33	18	5	0	0	0	0	0	0	0
137	75	28	154	240	215	151	89	53	22	6	0	0	0	0	0	0	0
138	242	235	149	75	110	251	151	90	35	11	5	1	1	0	0	0	0
139	59	86	211	313	276	79	208	145	113	64	17	8	1	1	0	0	0
140	185	173	119	179	314	306	37	25	85	77	36	10	3	1	1	0	0
141	225	221	219	207	51	74	305	191	88	33	54	25	17	7	2	1	0
142	20	39	57	112	351	332	58	212	189	164	82	46	26	10	8	1	0
143	154	207	249	218	103	371	299	89	139	207	127	73	63	17	7	4	1
144	107	161	210	277	308	41	40	218	123	31	28	133	67	30	19	6	3
145	83	125	171	259	327	419	317	105	133	248	183	22	104	54	21	14	2
146	58	89	133	137	109	373	356	345	160	94	236	178	30	86	36	28	8
147	36	73	48	98	214	20	51	319	302	229	62	193	157	13	36	25	16
148	31	55	109	203	293	362	331	74	12	113	302	29	22	101	24	8	8
149	16	38	76	132	200	341	376	366	373	259	37	269	200	31	76	42	22
150	16	20	109	158	154	290	342	339	335	359	312	20	255	154	31	35	33
151	8	20	33	65	137	208	268	367	352	362	386	330	17	10	76	33	26
152	7	10	25	47	108	190	262	285	132	20	55	380	289	181	41	42	44
153	4	8	11	31	65	136	193	240	286	389	378	39	26	250	141	67	29
154	5	6	8	22	54	109	151	224	248	370	388	400	337	44	139	124	72
155	2	1	5	14	30	80	102	176	211	308	400	419	353	284	59	88	110
156	2	2	4	14	21	36	82	137	202	283	414	434	452	311	256	85	41
157	0	1	3	8	22	39	77	103	244	427	328	372	456	366	263	170	142
158	0	1	1	1	5	24	48	74	121	167	273	377	426	341	294	205	171
159	0	0	1	2	7	30	33	64	81	169	246	335	423	390	320	241	192
160	0	1	0	1	5	9	15	33	60	123	208	296	403	388	323	253	212
161	0	0	0	0	0	6	22	23	37	92	162	267	378	422	344	242	237

Table 8.9: Bivariate frequency between PSAE Science and ISBE-developed Science

(ISBE-developed science scores 58–74)

	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
162	0	0	0	1	0	4	6	19	35	82	214	217	294	367	390	268	252
163	0	0	0	0	0	2	3	17	26	33	80	161	239	303	329	272	286
164	0	0	0	0	0	2	3	4	13	28	51	164	218	258	324	271	264
165	0	0	0	0	0	2	1	0	12	21	60	176	184	233	279	274	271
166	0	0	0	0	0	0	1	0	2	10	29	62	286	216	269	251	258
167	0	0	0	0	0	0	0	0	0	5	21	38	77	165	224	249	241
168	0	0	0	0	0	0	0	0	1	8	17	29	74	245	184	195	206
169	0	0	0	0	0	0	0	0	3	1	8	17	39	82	255	187	222
170	0	0	0	0	0	0	0	1	1	0	2	10	28	71	95	167	173
171	0	0	0	0	0	0	0	0	0	0	2	2	13	40	60	172	189
172	0	0	0	0	0	0	0	0	0	0	4	4	8	28	42	66	106
173	0	0	0	0	0	0	0	0	0	0	0	2	5	12	40	64	78
174	0	0	0	0	0	0	0	0	0	0	0	1	2	10	22	47	63
175	0	0	0	0	0	0	0	0	0	0	0	0	3	6	10	31	47
176	0	0	0	0	0	0	0	0	0	0	0	0	1	3	11	11	32
177	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	8	15
178	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	7
179	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6
180	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
181	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
182	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
183	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
184	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
185	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
187	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
188	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
189	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
190	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
191	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
192	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
193	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
194	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
195	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
196	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
197	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
198	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mn*	138	139	141	142	144	145	147	149	150	152	154	156	157	159	161	163	164

Note: N=116,960 (any examinee not having PSAE scores was deleted). *is rounded conditional mean given ISBE science scores.

Table 8.9: Bivariate frequency between PSAE Science and ISBE-developed Science

(ISBE-developed science scores 75–100)

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
125	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
126	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
127	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
128	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
129	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
130	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
131	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
133	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
134	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
135	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
136	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
137	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
138	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
139	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
140	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
141	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
142	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
143	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
144	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
145	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
146	5	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
147	17	6	4	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
148	15	10	5	4	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
149	13	11	7	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
150	28	8	9	4	2	2	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
151	31	26	10	7	6	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
152	18	15	19	14	1	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
153	50	19	9	9	7	8	2	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
154	29	35	25	7	9	3	5	3	2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
155	74	33	21	7	8	8	3	2	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
156	69	66	39	13	18	8	3	3	3	3	0	0	2	1	0	1	0	0	0	0	0	0	0	0	0	0
157	38	13	19	15	8	8	7	2	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
158	116	89	47	23	13	4	4	6	2	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
159	166	123	75	40	17	12	4	3	3	2	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0
160	197	173	105	58	22	13	9	8	3	2	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
161	212	205	125	78	37	30	22	9	6	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
162	229	100	74	101	41	41	20	12	14	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
163	267	245	110	57	47	28	22	13	9	1	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 8.9: Bivariate frequency between PSAE Science and ISBE-developed Science

(ISBE-developed science scores 75–100)

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
164	263	264	171	109	30	9	7	12	17	10	6	3	5	0	0	0	0	0	0	0	0	0	0	0	0	0
165	254	438	325	176	98	65	36	24	11	10	3	7	0	4	0	0	0	0	0	0	0	0	0	0	0	0
166	261	310	361	291	121	100	52	39	23	2	3	3	2	2	1	1	0	0	0	0	0	0	0	0	0	0
167	261	350	284	298	177	136	108	60	33	18	6	8	2	3	0	1	0	1	0	0	0	0	0	0	0	0
168	276	323	273	214	160	129	78	65	43	32	13	6	6	1	0	0	0	1	0	0	0	0	0	0	0	0
169	219	305	274	235	177	201	142	151	111	59	25	22	8	5	3	0	0	0	0	0	0	0	0	0	0	0
170	208	272	304	288	152	156	128	114	85	71	35	24	13	7	3	2	0	0	0	0	0	0	0	0	0	0
171	177	256	262	268	189	163	117	102	79	30	20	11	15	8	5	4	1	1	0	0	0	0	0	0	0	0
172	261	280	248	263	200	178	151	131	125	93	89	55	38	12	6	2	0	0	1	0	0	0	0	0	0	0
173	96	211	238	225	179	174	135	162	105	41	33	19	26	11	6	3	1	3	0	0	0	0	0	0	0	0
174	79	261	277	251	210	215	160	139	150	150	92	73	51	27	21	22	3	8	2	1	0	1	0	0	0	0
175	61	100	252	316	182	207	122	155	130	93	98	70	49	29	18	18	4	6	4	0	2	0	0	0	0	0
176	42	86	121	228	293	235	142	208	169	117	98	78	76	28	20	19	4	7	2	1	2	0	0	0	0	0
177	33	76	97	132	117	268	214	231	191	122	113	77	71	39	22	27	6	11	2	0	1	0	0	0	0	0
178	18	57	64	118	89	146	151	293	232	179	126	97	84	46	28	24	9	15	7	1	3	0	0	0	0	0
179	9	27	49	87	95	138	109	157	153	133	129	91	99	52	40	37	7	13	9	2	3	1	0	1	0	0
180	12	19	42	75	70	101	102	170	175	193	237	173	170	78	62	38	12	28	11	6	2	2	2	1	0	0
181	1	7	26	54	41	93	56	124	112	128	129	82	171	116	99	50	13	26	9	2	10	0	4	0	0	0
182	0	1	15	34	32	48	54	84	96	110	120	92	116	83	84	110	37	40	8	6	5	1	2	1	0	0
183	0	0	0	21	25	40	41	70	97	97	118	88	121	81	68	74	25	86	33	9	6	0	3	1	0	0
184	0	0	0	1	7	30	17	58	54	87	104	89	120	80	72	63	27	51	13	24	31	3	6	1	0	1
185	0	0	0	0	0	4	9	50	49	53	86	84	114	60	78	89	23	32	22	13	21	1	8	3	0	1
186	0	0	0	0	0	0	1	4	25	50	77	70	104	62	82	70	27	67	37	25	35	5	4	3	0	2
187	0	0	0	0	0	0	0	0	1	8	59	47	114	57	61	82	29	68	34	18	23	2	14	2	0	5
188	0	0	0	0	0	0	0	0	0	0	5	3	47	41	67	74	24	71	31	28	39	2	15	2	0	0
189	0	0	0	0	0	0	0	0	0	0	0	0	29	35	55	66	35	70	34	0	3	0	0	1	0	4
190	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	31	10	2	0	23	27	1	26	6	0	3
191	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	16	15	57	35	35	46	5	22	6	0	0
192	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
193	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	76	24	44	41	3	34	9	0	0
194	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	33	23	2	0	0	0	2
195	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	20	0	0	0	0	0
196	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	37	3	0	9
197	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	7	0	0
198	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	3	0	13
199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	16
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	9
Mn*	165	167	169	170	171	173	173	175	176	177	179	179	181	181	183	183	185	186	186	189	188	187	192	191	NA	195

Note: N=116,960 (any examinee not having PSAE scores was deleted). *is rounded conditional mean given ISBE science scores.

Table 8.10 presents the correlations among the three PSAE scores. The correlations are fairly homogenous with an average value of about 0.83 and a range of about 0.79 to 0.86. This homogeneity among the correlations suggests that one component can explain most of the variance among the three tests. Tables 8.11 and 8.12 present the results of a principal component analysis of the correlation matrix for the three tests. Table 8.11 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 a little more than 88% of the variance among the three tests. The remaining components all have eigenvalues less than one, and both combined only account for about 12% of the variability. This further indicates a one-component model fits the data well. Table 8.12 contains the loadings of the three tests on the first principal component across years for the PSAE. 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, and provides evidence of consistency of the PSAE construct across years.

The correlational analysis and the principal component analysis among the various PSAE subject scores reflect on the internal structure of the total PSAE test battery. Though the number of test subjects comprising the battery has differed over time, the results of the aforementioned analyses have been very consistent over time, and this indicates that the internal structure of the test battery has been very stable from year to year.

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

Table 8.10: Correlations among PSAE scores

	Reading	Mathematics	Science
Reading	1.00000	0.78649	0.82853
Mathematics	0.78649	1.00000	0.85803
Science	0.82853	0.85803	1.00000

Table 8.11: Eigenvalues of the Correlation Matrix

Component	Eigenvalue	Difference	Proportion	Cumulative
1	2.64904299	2.43199973	0.8830	0.8830
2	0.21704326	0.08312950	0.0723	0.9554
3	0.13391375		0.0446	1.0000

Table 8.12: First Principle Component loading values across years

PSAE Area	First Principle Component Loadings					
	2001	2002	2003	2004	2005	2006
Reading	.91	0.92	0.92	0.92	0.93	0.93
Mathematics	.91	0.91	0.91	0.91	0.94	0.94
Science	.94	0.95	0.95	0.94	0.96	0.95

Figure 8.1: Percentage of all students who meet the Illinois Learning Standards in 0, 1, 2, or 3 PSAE subjects for PSAE Spring 2006

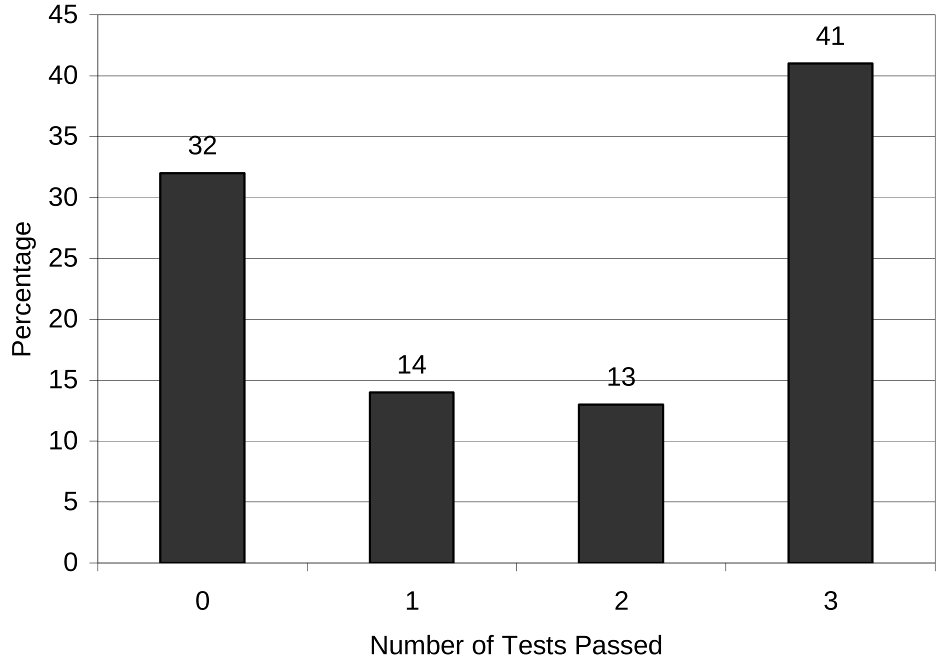


Figure 8.2: Percentage of students who meet the Illinois Learning Standards in 0, 1, 2, or 3 PSAE subjects for PSAE Spring 2006 – By gender

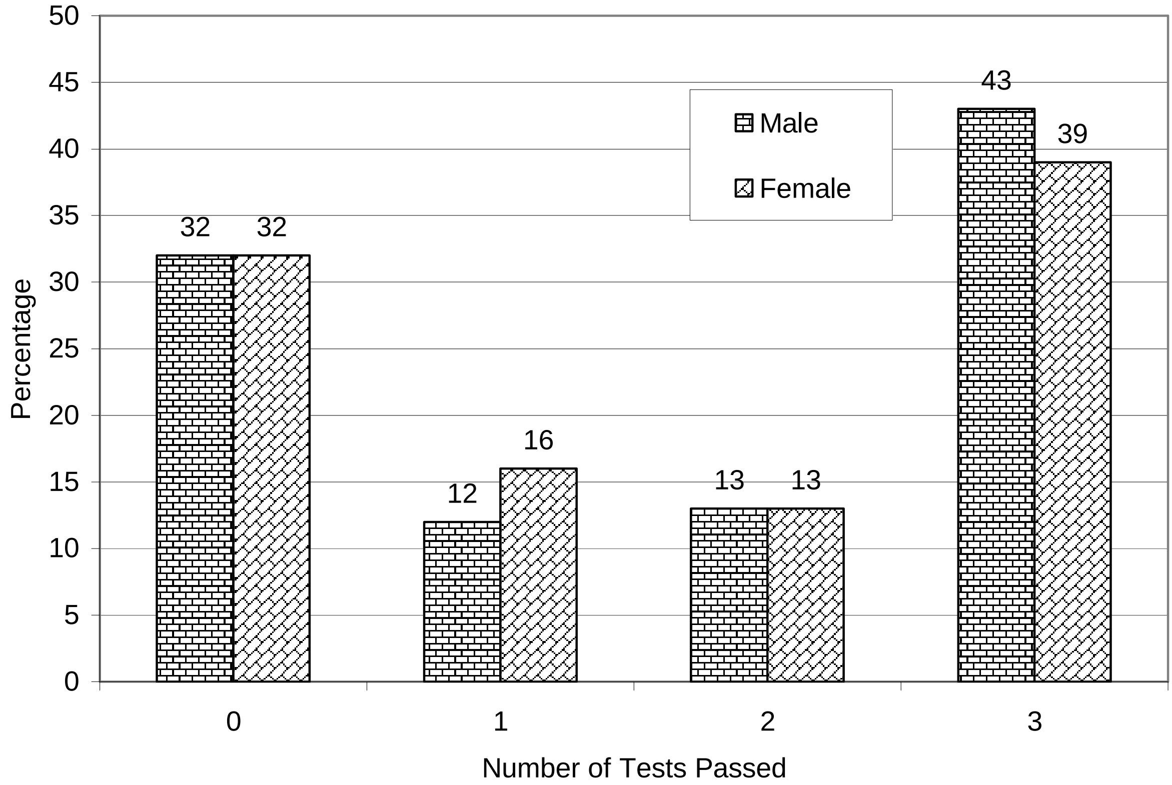
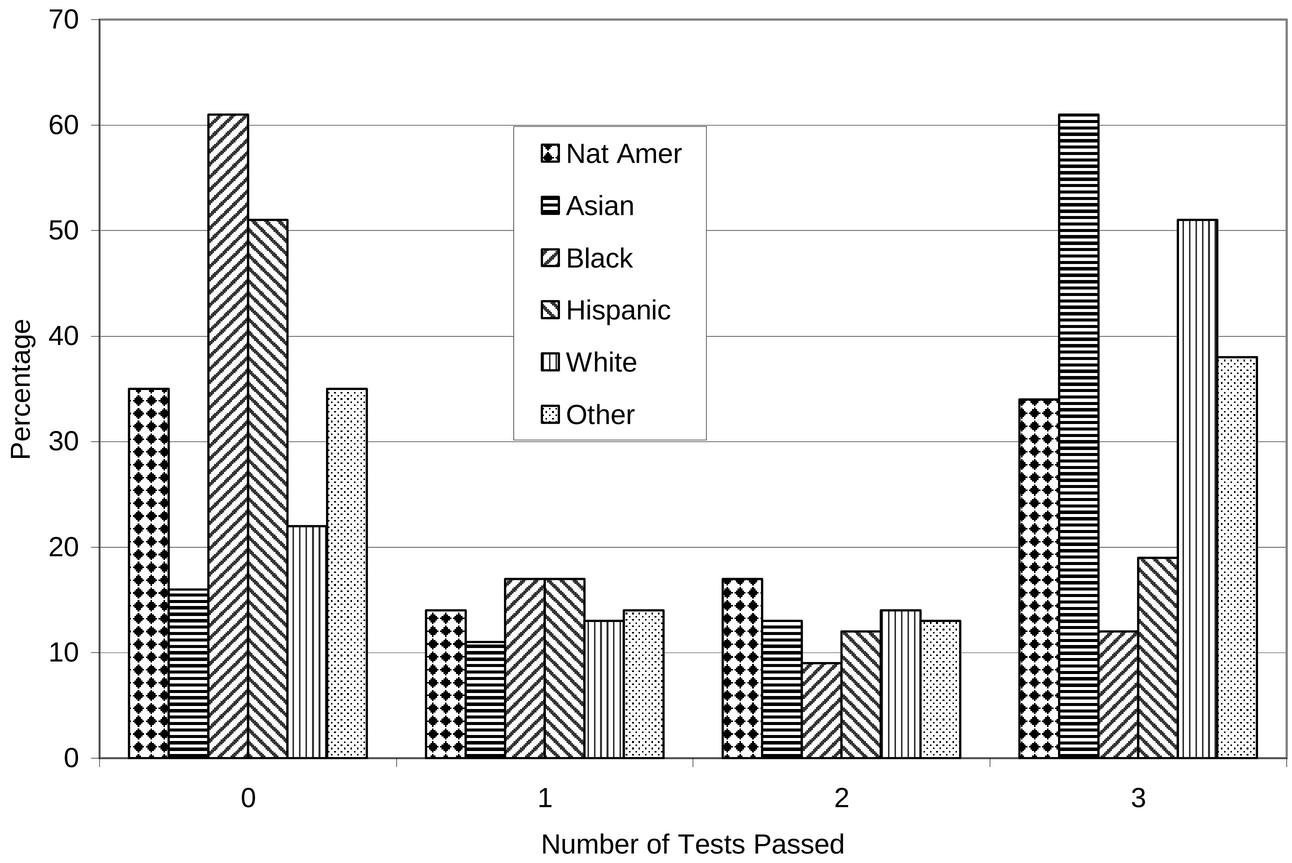


Figure 8.3: Percentage of students who meet the Illinois Learning Standards in 0, 1, 2, or 3 PSAE subjects for PSAE Spring 2006 – By race/ethnicity



Chapter 9
Fall 2006 PSAE Grade 12 Retake

In fall 2006, ISBE offered grade 12 Illinois public school students the option of retaking the PSAE or of taking it for the first time if they had not taken it during the spring 2006 administration. The retake option was mandated by Illinois law to provide students with an opportunity to raise their PSAE scores and to allow students who did not participate in the grade 11 administration to earn a set of PSAE scores. The fall administration does *not* count towards school accountability. More than 6,000 students took advantage of the 2006 retake option.

Day 1 testing was offered on a national ACT test date: Saturday, October 28, 2006. Students took the test at national test centers. Day 2 testing was offered in schools on the Tuesday following Day 1 testing: October 31, 2006. Students were required to take both the Day 1 and Day 2 tests in October to earn new PSAE scores in reading, mathematics, and science.

If a student earns PSAE scores in both the spring and fall administrations, only the higher PSAE score in each subject is reported on his or her permanent record. For example, a student might receive the higher PSAE score in reading during the April 2006 administrations and the higher PSAE mathematics score during the October 2006 administration; in that case the April reading score and the October mathematics score would be recorded on the student's permanent record. The data set that includes only the higher scores in each subject for students who took both the spring 2006 and fall 2006 PSAE is defined as the *best* group.

Table 9.1 shows the percentage of students in each of the PSAE performance levels by test and by test administration. The student scores included in Table 9.1 are for those who took the PSAE both in spring 2006 and fall 2006; at the group level, these students improved their PSAE scores. The percentage of students meeting or exceeding the Illinois Learning Standards is highest for the best group. It is not known whether the increase is due to practice—that is, students become more familiar with the test format—or due to higher performing students being more likely to repeat the test.

Table 9.2 shows the differences between the spring and fall averages for both the PSAE scores and the component tests, the standard deviations of these differences, and the minimum and maximum change observed across all students.

Table 9.3 provides the percentage of examinees in each performance level for those students who took the PSAE in fall 2006. Table 9.3 includes all students with fall 2006 scores, whether they had spring 2006 scores or not. Table 9.4 provides the moments for the PSAE score for all examinees who received PSAE scores from the fall 2006 administration.

Table 9.1: Percentage of students in each PSAE performance level by test and by test administration for students who participated in both the spring and fall 2006 administrations

Subject	Performance Level	Spring 2006	Fall 2006	Best[a]
Reading	Academic Warning	5.31	4.38	3.83
	Below	40.25	39.72	36.65
	Meets	47.61	48.10	50.96
	Exceeds	6.83	7.80	8.56
	No. of missing records	356	2,462	218
Mathematics	Academic Warning	7.90	4.55	5.69
	Below	45.49	46.46	43.94
	Meets	43.53	43.91	45.82
	Exceeds	3.08	5.08	4.55
	No. of missing records	350	2,461	215
Science	Academic Warning	6.82	6.32	4.92
	Below	54.36	48.10	50.32
	Meets	34.10	39.11	38.21
	Exceeds	4.71	6.47	6.55
	No. of missing records	352	2,462	217

[a] This column reflects the best performance of students who earned PSAE scores in both spring *and* fall administrations.

Table 9.2: Means of difference in PSAE scores between the spring 2006 and fall 2006 administrations

Variable	N	Mean	Std Dev	Minimum	Maximum
PSAE Reading	3,309	–0.27	9.19	–50	46
PSAE Mathematics	3,313	0.56	6.46	–30	37
PSAE Science	3,312	0.86	7.88	–33	29
WorkKeys Mathematics	4,127	0.11	1.19	–7	6
WorkKeys Reading	4,125	–0.16	1.21	–7	7
ACT Math	4,905	0.44	2.16	–13	17
ACT Reading	4,899	0.82	3.65	–17	20
ACT Science	4,899	0.72	3.37	–16	16
ISBE Science	4,128	–0.25	6.09	–30	25

Note: Difference is fall minus spring; thus a positive mean reflects improvement.

Table 9.3: Percentage per performance level for fall 2006 PSAE

	Academic Warning	Below	Meets	Exceeds
Reading	7.12	43.51	42.93	6.43
Mathematics	8.01	50.85	37.13	4.01
Science	10.64	50.68	33.70	4.97

Table 9.4: Fall 2006 PSAE composite moments

Variable	N	Means	Std Dev	Minimum	Maximum	Skewness	Kurtosis
Reading	5488	154.88	14.19	120	195	0.01	–0.14
Mathematics	5491	153.25	13.45	120	200	0.33	0.35
Science	5489	153.50	14.08	120	197	0.17	–0.58

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. (2001). *WorkKeys technical handbook*. Iowa City, IA: Author.

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

ACT. (2007a). *ACT National Curriculum Survey 2005–2006*. Iowa City, IA: Author.

AERA. See American Education 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.

Brennan, R. L., & Lee, W. (1997). *Conditional standard errors of measurement for scale scores using binomial and compound binomial assumptions*. Iowa Testing Programs Occasional Paper Number 41. Iowa City, IA: University of Iowa.

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

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

Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement for scale scores. *Journal of Educational Measurement, 29,* 285–307.

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., 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. (1955). A survey of observed test-score distributions with respect to skewness and kurtosis. *Educational and Psychological Measurement, 15,* 383–389.

Lord, F. M. (1957). Do tests of the same length have the same standard errors of measurement? *Educational and Psychological Measurement*, *17,* 510–521.

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.

Webb, N. L. (2006, 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. (2006, 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. (2006, 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 Assessment Test Accommodations
for Day 1 of the Prairie State Achievement Examination

PROCEDURES FOR REQUESTING ACT ASSESSMENT® TEST ACCOMMODATIONS
for Day 1 of the Prairie State Achievement Examination (PSAE)
(April 2001)

REQUESTING TEST ACCOMMODATIONS

By direction of the Illinois State Board of Education, requests for accommodations on the ACT Assessment administered as part of the PSAE may be submitted **only** for students currently receiving special education and related services described in an Individualized Education Program (IEP) developed in accordance with Section 614 of the Individuals with Disabilities Education Act (IDEA) or students with current 504 Plans.

Accommodations are not available for students with limited English proficiency (LEP).

ELIGIBILITY REQUIREMENTS

To be considered for test accommodations, students must meet **ALL** of the following ACT Assessment 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" on page 2.
   - If diagnosed for the FIRST time before September 1997, reconfirmation is required within the last 3 years. A current IEP or 504 Plan on file at the public school may serve as reconfirmation, provided the initial diagnosis was made by a qualified professional(s)
   - If FIRST diagnosed within the last 12 months, full written diagnostic documentation must be submitted *with* the request form.
2. Current IEP or 504 Plan. The request must be submitted with a copy of the test accommodations page from the student's *current* IEP or a complete copy of the *current* 504 Plan that supports the need for all requested accommodations due to the disability.
3. Extended Time and Test Accommodations Allowed in School. The student must currently receive requested accommodations, including extended time, for tests in school due to the disability.

AUTHORIZATION TO PROVIDE CONFIDENTIAL DOCUMENTATION

Schools are required to provide the necessary information and documentation to support requests for test accommodations. The Illinois State Board of Education has authorized ACT to collect and review this documentation.

All documentation provided to ACT will be kept confidential, will be used solely to determine eligibility, and will not become part of the student's ACT Assessment score record.

OPTIONS FOR TEST ACCOMMODATIONS

Examples of test accommodations include the following:

1. Standard Print and Standard Time. If a student can test in a single session with standard time limits (including standard break) and use a standard (10-point) test booklet, but the disability requires other accommodations, *the school may make arrangements for such accommodations without prior consultation with ACT.* Examples include, but are not limited to:
   - assignment to a wheelchair accessible room,
   - permission to eat snacks if diabetic,
   - permission to use Irlen filters or color overlays.

   Examples of accommodations for students with hearing impairments that do not require time extensions 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 lip-read spoken instructions,
   - a written copy of spoken instructions with visual notification from testing staff of test start, five minutes remaining, and stop times.
2. Large Type and Standard Time. If a student can test with standard time limits (including standard break) but needs a large-type (18-point) test booklet, the school must submit a completed request form specifying the accommodations requested.
3. Extended Time and/or Alternate Formats. If the student's professionally diagnosed and documented disability requires one of the accommodations below, the school must submit a completed request form:
   - More than standard time
   - Testing over more than one day
   - Stop-the-clock breaks
   - Alternate test formats such as braille, audio cassette, or a reader.

DEADLINE FOR SUBMITTING REQUESTS

To be considered for the April 25, 2001, Day 1 PSAE test date, requests and all required documentation must be **received by ACT no later than February 1, 2001**. The Test Accommodations Coordinator at each school is responsible for gathering all completed requests and submitting them as a group *with* a completed Test Accommodations Coordinator Header to:

ACT State Test Accommodations—(61)
2255 North Dubuque Road
PO Box 4071
Iowa City, IA 52243-4071

Phone: 319/337-1332 (voice)
319/337-1701 (TDD)
8:30 A.M.–4:30 P.M., Central Time, M-F

INSTRUCTIONS FOR COMPLETING THE REQUEST FORM

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 request form will NOT be processed if it is:

- incomplete or missing required documentation,
- not signed by both student/parent **and** school official,
- not submitted with a Test Accommodations Coordinator Header, or
- not received by the deadline.

SIDE 1

A. Student Information. Print or type. Student address is especially important for students who plan to take the ACT Assessment again.

B. Diagnosis/Disability. Check all applicable disability diagnoses as stated in written documentation on file at the school.

C. Test Format Requested. Documentation of a visual disability is required to support requests for large-type test booklets. (Both regular-type and large-block answer sheets are provided with each large-type booklet.)

D. Time Requested. Mark the option closest to the accommodations normally provided at school.

E. Other Accommodations Requested. If other accommodations are needed due to the disability, explain in detail and submit supporting documentation.

SIDE 2

F. Diagnosis. Provide ALL requested information.

F-1. List the specific diagnosis as stated in the written documentation on file at the school.

F-2. The credentials of the diagnosing professional, or team of professionals, must be appropriate to the disability. If the disability was identified by an IEP team, list relevant titles and specializations.

F-2a. Specify the date of the **FIRST** diagnosis, usually early in a student's education. **If the FIRST diagnosis was within the last 12 months, submit complete diagnostic documentation *with* the request form** (see "Guidelines for Documentation" on this page).

F-2b. **If FIRST diagnosed before September 1997, there must be a reconfirmation 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 public school may serve as reconfirmation.

G. Current IEP or 504 Plan on File at School.

G-1. Indicate the type of plan currently on file at the school and attach the required copy. (The student's name and effective dates of the plan must appear on the pages submitted.)

G-2. Mark every year during which the student has had an IEP or 504 Plan documenting the need for the requested accommodations.

H. Current Time Accommodations at School. Indicate the time normally permitted for this student for classroom and standardized tests as a result of this diagnosis and supporting documentation.

I. Previous Approval of Extended Time on ACT Assessment. Mark the appropriate answer. If yes, provision of the student's Social Security Number will speed processing of this request.

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

K. Student/Parent Signature. Depending on the student's age, the student or parent must read and sign the statement. *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.*

ACT GUIDELINES FOR DOCUMENTATION

Documentation, which must be on file at the school, must be written by the diagnosing professional(s) and must meet **ALL** of these guidelines:

1. **States the specific disability** as diagnosed
2. **Is current** (no older than September 1997)
3. **Describes the 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*, provides test results, with standard scores and percentiles (including subtests), 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*, includes
     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 how they impair the individual.
   - *For visual, hearing, and other physical/psychological disorders*, provides results from complete ocular, audiologic, or other appropriate medical examination.
5. **Describes the functional limitations resulting from the disability**, as supported by the test results
6. **Describes specific recommended accommodations** and provides a rationale explaining how these specific accommodations address the functional limitations
7. **Establishes the professional credentials of the evaluator**, including information about licensure or certification, education, and area of specialization

Additional details about ACT's policies for documentation of requests for testing accommodations are available on ACT's website at: **www.act.org**. Click on "ACT Assessment," then "Services for Students with Disabilities," then "ACT Policy for Documentation to Support Requests for Testing Accommodations on the ACT Assessment."

Related Information

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

1. Requesting Additional Information. If ACT has a question about the diagnosis or accommodations, the Test Accommodations Coordinator at the school will receive a request for information before ACT's review of the request can be completed. If ACT requests additional information, a fax reply will assist in meeting deadlines. Documentation must be submitted in writing, not by phone. If the request cannot be approved by April 18, 2001, the student will be allowed to test with standard time limits and use a regular print (10-point) test booklet without accommodations.

2. Notification of Day 1 Accommodations. ACT will send to the school's Test Accommodations Coordinator an authorized accommodations letter for each student approved for accommodations on the ACT Assessment during the PSAE administration. If a request is not approved, ACT will send written notification to the school.

3. Determining Day 2 Accommodations. ACT's approval of accommodations applies to Day 1 administration of the ACT Assessment **only**. School personnel are advised to use ACT's approval as a guideline for Day 2 accommodations to the extent that the same accommodations are appropriate to the nature of Day 2 tests and consistent with a student's IEP or 504 Plan. There is **no** request form for Day 2 accommodations.

4. Extended-Time Testing Windows. Testing for students approved for extended-time or multiple-day testing must be completed during the applicable two-week PSAE testing window:
   - April 25-May 9, 2001, for Day 1 (ACT Assessment)
   - April 26-May 10, 2001, for Day 2 tests

5. Assignment of Test Materials. ACT Assessment test materials may be used ONLY for each student approved for accommodations and may NOT be transferred to any other school or examinee.

6. ACT Assessment Score Reports. ACT Assessment scores achieved with extended time through PSAE testing will be reported as "State Special." ACT does not report any specific information about the disability or accommodations authorized.

7. Preparing for the ACT Assessment. *Preparing for the ACT Assessment* provides information about the tests and includes a complete practice test. Schools have a supply of this free regular-type booklet for distribution to students. Many schools have previously ordered copies of a practice test in braille, large-type, or on cassettes 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. (See phone number on page 1.)

ACT Assessment Repeat Testing

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

1. During Spring 2001. ACT has adjusted its usual 60-day ACT Assessment retest restriction for the PSAE administration. Grade 11 students who wish to take the ACT Assessment 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 during the PSAE testing window for ACT Assessment accommodations ***and*** may also apply for ACT Assessment Extended-Time National Testing on the April 7, 2001, national test date. Students must submit a complete Application for Extended-Time National Testing postmarked by March 2, 2001.
   - 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 during the PSAE testing window for ACT Assessment accommodations ***and*** may also apply for and test via ACT Assessment Special Testing during February and/or March 2001. Students must submit a complete Request for Special Testing a minimum of four weeks prior to their proposed Special Testing.

2. During 2001-2002. Students who have been approved for ACT Assessment accommodations for the spring PSAE administration and wish to retake the ACT Assessment with accommodations during 2001-2002 will be eligible for a streamlined request process next year.

   These students will first need to determine which of the following options is appropriate to their disabilities:
   - ACT Assessment 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 Assessment 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, audio cassette, or a reader).

   These students will need to submit the appropriate 2001-2002 ACT Assessment request form with only Side 1 completed, along with a copy of their authorized accommodations letter from the PSAE administration of the ACT Assessment. This process will eliminate the need for completing Side 2 of the form and will streamline review of the request in light of prior approval.

April 2001 Request For Day 1 PSAE Test Accommodations On The ACT Assessment®

This form is to be completed by a school official, such as counselor, special education teacher, or principal, following the instructions on page 2 of the folder entitled "Procedures for Requesting ACT Assessment Test Accommodations for Day 1 of the PSAE."

A. STUDENT INFORMATION. (Please print clearly.)

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

Student Street Address or PO Box ____ City ____ State ____ Zip ____

Name of High School Where Student Is Enrolled (may differ from school student attends) ____ ACT High School Code (required) ☐☐☐-☐☐☐

B. DIAGNOSIS/DISABILITY. (Check all that apply.)

Learning Disability (01)
- ☐ (DA) Developmental Arithmetic Disorder
- ☐ (DY) Developmental Reading Disorder (Dyslexia)
- ☐ (LD) Other Learning Disability (explain on side 2, F-1)

Cognitive Disability (03)
- ☐ (AD) Attention Deficit Disorder/ADHD

Psychological Disability (03) (**FULL DOCUMENTATION REQUIRED**)
- ☐ (AX) Anxiety Disorder (explain on side 2, F-1)
- ☐ (PD) Other Psychological/Cognitive Disability (explain on side 2, F-1)

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

Other Disability (07)
- ☐ (OD) Other (explain on side 2, F-1)

C. TEST FORMAT REQUESTED. (Mark at least one.)
- ☐ (31) Regular-Type (10-point)
- ☐ (32) Large-Type (18-point)
- ☐ (33) Braille (printed copy included)
- ☐ (34) Cassette with Regular-Type
- ☐ (35) Cassette with Large-Type
- ☐ (36) Cassette with Raised Line/Braille Tables and Illustrations
- ☐ (37) Reader's Script* with Regular-Type
- ☐ (38) Reader's Script* with Large-Type
- ☐ (39) Reader's Script* with Raised Line/Braille Tables and Illustrations

* Note: A student approved to test with a reader must test *individually*. Readers may *not* read the tests to a group of examinees.

D. TIME REQUESTED. (Mark the option most similar to that provided in school.)
- ☐ (01) Standard time in <u>one session with standard break</u>.
- ☐ (05) Standard time on each test; authorization to test over <u>multiple days</u>.
- ☐ (06) Up to 50 percent additional time (time-and-a-half): up to <u>5 hours total</u> to complete four tests in <u>one session</u> (including breaks between tests)
- ☐ (02) Up to 100 percent additional time (double time): up to <u>7 hours total</u> to complete four tests in <u>one day</u> (including breaks between tests and a break for lunch)
- ☐ (04) Extended time on each test; authorization to test over <u>multiple days</u>.

E. OTHER ACCOMMODATIONS REQUESTED. (Mark only if applicable **and enclose supporting documentation**.) Students approved for extended time will test in a different room from the standard-time administration. It is your responsibility to indicate other accommodations that are needed **in addition to** extended time, and enclose supporting documentation.
- ☐ Written copy of spoken instructions
- ☐ Authorization to bring sign language interpreter (not a relative) for **spoken** instructions (not test items)
- ☐ Wheelchair access; table (not desk)
- ☐ Stop-the-clock breaks
- ☐ Other (be specific): ____________________________

SUBMITTING THE REQUEST. Incomplete or unsigned forms will delay processing, which may cause the student to miss testing with accommodations. The request must be submitted ***with*** a signed Test Accommodations Coordinator Header. Address all requests from your school as a group to: ACT State Test Accommodations (61), 2255 North Dubuque Road, PO Box 4071, Iowa City IA 52243-4071. All submissions must be **received** at ACT by **February 1, 2001**. Early applications are encouraged. If ACT has questions about the information submitted, the Test Accommodations Coordinator will be contacted by fax.

________________________________ Student's Name (please print)

________________________________ Social Security Number (optional)

F. DIAGNOSIS.

1. Specific disability as stated in documentation on file (must be more specific than "learning disability"):

____

2. When and by whom student was:

	a. **FIRST** diagnosed*.	b. re-diagnosed **(within last 3 yrs).**
Date (month/year):	__________	__________
Age or grade of student:	__________	__________
Person or team making diagnosis:		
Name/team	__________	__________
Job title(s)	__________	__________
Institutional affiliation	__________	__________
Qualifications (degrees, specialization, certification)	__________	__________

***COMPLETE DOCUMENTATION REQUIRED if FIRST diagnosed within last 12 months;** s**ee "Guidelines for Documentation."**

G. 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.

1. Mark the type of plan currently on file for this student and attach the required copy.
   - ☐ IEP; attach a copy of the test accommodations page from the most current IEP.
   - ☐ 504 Plan; attach a complete copy of the most current 504 Plan.

2. Mark every year for which the student has had an IEP or 504 Plan.
   ☐ 2000-2001 ☐ 1999-2000 ☐ 1998-1999 ☐ 1997-1998

H. CURRENT TIME ACCOMMODATIONS AT SCHOOL. (Mark all that apply.) Indicate the time accommodations normally permitted for this student for classroom and standardized tests as a result of this diagnosis and supporting documentation.

- ☐ About 50 percent additional time (time-and-a-half)
- ☐ About 100 percent additional time (double time)
- ☐ More than double time
- ☐ Testing over more than one day
- ☐ Other (explain):________________________________

____

I. PREVIOUS APPROVAL OF EXTENDED TIME ON ACT ASSESSMENT.

☐ Yes ☐ No

Has this student been approved previously for testing with extended time on the ACT Assessment?

(If yes, provision of the student's Social Security Number at the top of this page will allow ACT to verify that approval and speed processing of this request.)

J. SCHOOL OFFICIAL'S SIGNATURE. *I affirm the student named on this form is enrolled at this school (may differ from the school student attends), and I have completed this form based on the documentation on file at the school. I verify the information provided on this form* ***and in the attached IEP or 504 Plan and other required documentation*** *is accurate, to the best of my knowledge, and reflects the test accommodations now being provided in school.*

________________________________ School Official's Signature (may not be a relative of the student)

________________________________ Print Official's Name and Title

K. STUDENT/PARENT SIGNATURE. (Form cannot be processed without appropriate signature.) *I authorize release to ACT of full diagnostic information by school officials, physicians, or others having such information. I understand that any documentation provided to ACT will be kept confidential, will be used solely to determine eligibility, and will not become part of the student's permanent score record. I understand ACT Assessment scores earned over multiple days or with extended time will be reported as "State Special" on score reports. 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)

__________ Date

Appendix B

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

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 <u>and</u> 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 not 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 www.act.org. Students are responsible for supplying their own calculators; schools may, if they wish, lend calculators to students who need to borrow one.
>
> A formula sheet is provided as part of the test booklet for the WorkKeys *Applied Mathematics* assessment. However, students are not 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

<table>
<tr><th></th><th>58B</th><th>58E</th><th>57B</th><th>ACT Average</th><th>NAEP</th><th>A07BB</th><th>C10BB</th><th>Example</th><th>WorkKeys Average</th><th>NAEP</th><th>Merged Mean</th><th>Within Range</th></tr>
<tr><td>Number</td><td>13</td><td>17</td><td>15</td><td>15</td><td>10</td><td>67</td><td>61</td><td>50</td><td>60</td><td>10</td><td>31</td><td>YES</td></tr>
<tr><td>Measurement</td><td>13</td><td>12</td><td>11</td><td>12</td><td rowspan="2">30</td><td>27</td><td>30</td><td>43</td><td>33</td><td rowspan="2">30</td><td>19</td><td>YES</td></tr>
<tr><td>Geometry Concepts Relations</td><td>19</td><td>21</td><td>25</td><td>22</td><td>0</td><td>0</td><td>0</td><td>0</td><td>14</td><td>YES</td></tr>
<tr><td>Algebra</td><td>45</td><td>40</td><td>45</td><td>43</td><td>35</td><td>0</td><td>0</td><td>0</td><td>0</td><td>35</td><td>28</td><td>YES</td></tr>
<tr><td>Data/chance</td><td>10</td><td>8</td><td>3</td><td>7</td><td>25</td><td>6</td><td>9</td><td>7</td><td>7</td><td>25</td><td>7</td><td>NO</td></tr>
</table>

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.