An Overview of the State Assessment System

Introduction

Education reform legislation, passed in 1985, began a process that started with adoption of 34 State Goals for Learning in eight learning areas and led 12 years later to the adoption of the Illinois Learning Standards. The State Goals are broadly stated expressions of the knowledge and skills that students should know and be able to do by the time they complete their secondary education. The Learning Standards describe in specific detail what students should know and be able to do in these learning areas.

Since the Illinois State Board of Education (ISBE) adopted the Standards, it has taken many steps towards helping schools provide all students the opportunity of meeting the Standards. To measure the progress that schools and students have made toward meeting the Standards, ISBE relies on four major state tests: the Illinois Standards Achievement Test (ISAT) for elementary and middle school students (and for high school students, on a voluntary basis for certain subjects); the Prairie State Achievement Examination (PSAE) for high school students; the Illinois Measure of Annual Growth in English (IMAGE) for students with limited English proficiency; and the Illinois Alternate Assessment (IAA) for students with disabilities for whom the ISAT or PSAE is inappropriate. The Standards are for all students, and all students are expected to make yearly progress in meeting the Standards.

The state assessment system, however, is only one of the tools that schools have available for planning and aligning their curricula to the Standards. It is not designed primarily as a diagnostic test for individual students. In fact, a single test cannot be used for diagnostic purposes. True diagnosis, whether it is in education or the medical field, requires many types of assessment and data analysis to determine what is needed to improve academic or physical health, respectively. The state assessment system is designed to provide accountability at the school, district, and state levels and to guide the program-development efforts of schools and districts. It provides reliable data that can be compared from year to year, which schools can use to assess the effectiveness of their curricular modifications.

By themselves, state tests cannot provide complete information about an individual student’s achievement. Any student can have a bad day, and not all information that students have mastered will be measured in a given year. Complete evaluation of an individual student must rely on evaluation of their day-to-day work in class, grading their homework assignments and classroom assessments, and keeping track of how they progress throughout the school year. Nevertheless, a generally accurate picture of achievement trends in schools, districts, and the state emerges over time when the results are reported for tests that are taken by nearly all public school students in the grades being tested. The resulting data give students, parents, and educators good information about how schools, districts, and the state are progressing with respect to the Standards. The tests also generate useful information about individual student achievement when they are considered in the context of the rest of a student’s academic record.

History and Update of the State Testing System

In 1983 and 1985, ISBE used the Illinois Inventory of Educational Progress (IIEP) pilot to gather information about students’ writing, mathematics, and reading achievement. Under 1985 reform legislation, the State Board instituted the Illinois Goal Assessment Program (IGAP) to provide a common yardstick for monitoring school and district progress in achieving the State Goals. The 1985 reform legislation was followed by accreditation system legislation and a school-improvement plan...
IGAP rules specified the activities and responsibilities of Illinois school districts and ISBE with respect to goals, objectives, plans, assessment, reporting, and improvement.

**Assessing the State Goals: the Illinois Goal Assessment Program (IGAP)**

IGAP testing began in 1988, with ISBE reporting only school and district data. The legislation that established IGAP mandated that seven of the eight State Goal learning areas be tested (reading, mathematics, writing, science, social science, physical development and health, and fine arts but not foreign language) at grades 3, 6, 8, and 11. Students took IGAP reading at grades 3, 6, and 8 in 1988. IGAP mathematics was added in 1989, IGAP writing in 1990, and IGAP science in 1991.

Before all subjects were folded into IGAP, however, the Superintendent’s Committee on Testing (chaired by a district superintendent and representing many educational partners) recommended major changes. As a result, instead of requiring testing of all subjects at every grade being tested, the General Assembly passed legislation that modified IGAP. In 1993, under the new schedule, IGAP assessed reading, mathematics, and writing at grades 3, 6, 8, and 10 and science and social science at grades 4, 7, and 11 and conducted special studies in physical development and health and in fine arts. Under this legislation, the General Assembly also limited state testing during the school year to 25 hours and required ISBE to report individual scores, beginning in 1994.

**Assessing the Illinois Learning Standards: the Illinois Standards Achievement Test (ISAT)**

Following adoption of the Illinois Learning Standards in 1997, ISBE developed the Illinois Standards Achievement Test (ISAT), and schools administered the first ISAT reading, mathematics, and writing assessments in 1999. The change from IGAP to ISAT testing took place with only one year of development and was accomplished without a break in testing. Additionally, the General Assembly mandated that reading, writing, and mathematics be tested at grade 5 under ISAT, rather than at grade 6 as under IGAP. Science and social science were added to ISAT in 2000. The State Board decided not to chart trends from IGAP to ISAT, so the 1999 ISAT established first-year baselines.

Schools administered ISAT for the third time in April 2001. Every major program in ISBE uses ISAT data. All grants and other applications for dollars require ISAT data. Title 1 funding is contingent on yearly progress demonstrated by ISAT scores. There are 31 test sessions, 14 of which include extended response, for almost 600,000 students in the state. Reports go back to schools showing how students, school, districts, and the state as a whole have performed with respect to the Illinois Learning Standards.

**Assessing the Illinois Learning Standards: the Prairie State Achievement Examination (PSAE)**

At the time that the switch from IGAP to ISAT took place, the General Assembly passed legislation that resulted in an additional grade 12 assessment, the Prairie State Achievement Examination, to be administered in January to grade 12 students and again to those students, if necessary, in March. Following the appointment of Superintendent M’Gee, new legislation was successfully obtained to consolidate the grades 10 and 11 ISATs with the grade 12 PSAE into one grade 11 PSAE.

On July 29, 1999, Governor George Ryan signed the PSAE legislation into law. Public Act 91-283 requires the PSAE to be administered to eligible grade 11 students beginning in spring 2001. The law also requires the PSAE to test reading, writing, mathematics, science, and social science and to be aligned with the Illinois Learning Standards.

The PSAE comprises the ACT Assessment®, two WorkKeys® tests, and three ISBE-developed tests. The ACT Assessment includes four curriculum-based tests that measure educational development in English, mathematics, reading, and science reasoning, thus measuring knowledge and skills in four of the domains of the Illinois Learning Standards.
the five academic areas of the PSAE. WorkKeys addresses the “Applications of Learning” that is a part of the Standards for every academic area and also increases the range of abilities assessed by the PSAE. The ISBE-developed tests cover those Standards in writing, science, and social science, not covered by the ACT Assessment or WorkKeys.

Report on the First Administration of the PSAE

On April 25 and 26, 2001, grade 11 students took the PSAE for the first time. The initial reports have been quite positive: Attendance was high, and students approached the test with a positive, serious attitude. School administrators put in hours of careful planning that resulted in a well-organized test administration. ISBE will be looking at survey data to analyze improvements for next year’s administration.

For ISBE staff, work on the Prairie State Achievement Examination (PSAE) began with the development and release of a Request for Sealed Proposals (RFSP) on April 4, 2000, to solicit bids for the services necessary to develop a PSAE system. A review of responses received by the May 4, 2000, deadline resulted in a recommendation to accept the proposal from ACT, Inc. Beginning early in June, ISBE and ACT staff engaged in a series of discussions and meetings to clarify and resolve the many outstanding issues that had been identified and put in writing by ISBE staff. On June 19, 2000, a final contract was awarded and signed.

During the next several months, ISBE staff completed the following tasks:

- Met with ACT and NCS Pearson (formerly National Computer Systems) staff to plan and work on details related to implementing the PSAE in 2001;
- Negotiated and finalized the purchase of ACT Assessment and WorkKeys test-preparation materials, which were shipped to schools in September 2000;
- Developed the 2001 PSAE Teachers’ Handbook and 2001 PSAE student-test preparation materials for distribution to schools;
- Developed materials for district and school administrator overview workshops, including a question-and-answer document and PowerPoint presentation;
- Arranged and conducted half-day overview workshops for district and school administrators from September 18 – 28, 2000, at 11 locations around the state;
- Met with the PSAE Advisory Committee to seek input on and discuss outstanding issues related to the first administration in 2001;
- Developed and disseminated a School Schedule and Site Options document to assist schools in meeting the PSAE requirements for standardized testing conditions; and
- Worked with ACT staff to finalize the procedures for schools to request test accommodations for students with disabilities on the ACT Assessment portion of the PSAE.

Later in the year, ISBE staff developed the 2001 PSAE Day 2 Supervisor’s Manual of Instructions and a question-and-answer brochure targeted to parents of Illinois students. Staff also collaborated with ACT to plan and conduct a series of 16 half-day PSAE Test Supervisor Training Workshops held throughout the state from February 26 – March 9, 2001. An audience of more than 1,400 test supervisors, back-up test supervisors, and ROE staff attended these trainings. In addition, ISBE staff responded daily to phone calls, faxes, e-mails, and letters asking all manner of questions about the PSAE.

These efforts contributed to a relatively smooth first administration of the PSAE in approximately 650 high schools on April 25 and 26, 2001.
Update on the Illinois Measure of Annual Growth in English (IMAGE) and LEP mathematics assessment

IMAGE is administered annually to students in grades 3 through 11 who do not take the ISAT or PSAE because of their lack of English-language skills. IMAGE measures reading and writing proficiency in English at three grade-level spans: grades 3 through 5, 6 through 8, and 9 through 11. Both the reading and writing tests are given in two 40-minute sessions, with students taking the test appropriate to their grade level. IMAGE measures the progress of students with limited English proficiency (LEP) in attaining the English-language reading and writing skills needed to meet and exceed the Illinois Learning Standards. The IMAGE reading and writing tests are administered annually to those students who have been enrolled in a state-approved bilingual education program (Transitional Bilingual Education or Transitional Program of Instruction) for less than three years, as an alternative for students to taking either the PSAE or ISAT because of their limited proficiency in English.

In 1999, ISBE staff and the IMAGE Advisory Committee developed a correlation between IMAGE and ISAT (IGAP at that time). The correlation yields predictive information on how students who take IMAGE would do if they were to take ISAT.

Last year 33,897 individual IMAGE student reports, which contain writing and reading scores for the student, school, district, and state, were generated. Students and their parents can see how they performed on the test compared to the school, district, and state. Teachers (for their classes) and administrators (for their schools) can see how students performed compared to each other, the school, the district, and the state. LEP students are permitted to take IMAGE for three years before taking the regular state assessment, so teachers and administrators can chart students’ progress toward English proficiency by comparing their scores from one year to the next. On average, LEP students improve their scores about 25 scale-score points per year in reading and about 2 points per year in writing.

Federal law requires that, at a minimum, states must assess LEP student performance in language arts and mathematics. To have a Title 1 assessment system eligible for U.S. Department of Education approval, ISBE had to add a mathematics assessment for LEP students who are not yet ready to take ISAT. For this reason, Assessment Division staff developed and pilot-tested a mathematics assessment to measure the Illinois Learning Standards in mathematics that accommodates the language learning needs of students.

This test was developed in close consultation with Illinois educators with expertise in mathematics and language acquisition. The pilot test data will be analyzed this summer with input from our contractor and Illinois educators to determine if this will be a valid and reliable assessment of the mathematics standards for these students.

Update on the first administration of the Illinois Alternate Assessment (IAA)

The 2000 – 2001 administration of the Illinois Alternate Assessment (IAA) was completed on April 13, 2001, which was the last day on which teachers could enter evidence in a student’s IAA portfolio. In contrast with other state assessments (which are standardized tests in multiple-choice and extended-response formats), the IAA has a portfolio format. Teachers collect evidence four times a year to show the progress a student is making to achieve the Standards-based goals in the student’s individualized education plan, or IEP, in a given content area. The portfolio may include student work samples, teacher observations, videos, photographs, charts, and other evidence. Schools mailed completed portfolios to the IAA Scoring Center in Bloomington, Illinois, from May 1 – 15, 2001. The contractor, Measured Progress, will oversee scoring of the IAA, which will take place from June through early August. Recruitment of a cadre of scorers, which will include Illinois special education teachers, is currently underway.

This year’s activities to disseminate information to educators and prepare them to implement the IAA included the following:
• From September 2000 through early April 2001, contractor and ISBE staff provided training on the IAA.

• On March 14, 2001, the Division of Intervention and Assessment conducted a live, interactive satellite teleconference that participants viewed at over 20 downlink sites across the state and via streaming Internet real video on the ISBE Web site.

• The document “Assessing Students with Disabilities: Questions and Answers” was distributed to schools, along with a cover memo explaining how the 2001 IAA results will be reported.

ISBE staff are currently working with Measured Progress to review the 2000 – 2001 IAA and develop recommendations for refinements to the process. These refinements will be based on input gathered over the past year from the IAA Advisory Task Force and the Ad Hoc Committee, as well as teachers, administrators, parents, and other stakeholders.

Prospective changes to the state assessment system: national proposals for every-year testing in grades 3 through 8

The position of the Bush administration and current legislative proposals include state testing of all students in grades 3 through 8 in reading and mathematics. At this time, only 13 states have state tests in all of these grades. Testing at these additional grades would require additional funding at both the national and state levels.

In addition to the increased testing as part of the Title 1 requirements – if legislation now pending in Congress is passed – the National Assessment of Educational Progress (NAEP) will be used to validate state assessments. The proposals for NAEP include testing every year at grades 4 and 8 in reading and mathematics. If Congress mandates this testing, the Illinois testing program would have to add at least grade 4 reading and mathematics to make direct comparisons.

Test Development: the Illinois Approach

The steps in developing ISAT, IMAGE, and the ISBE-developed tests of the PSAE are similar and follow procedures accepted by the industry for use in the development of standardized tests. The following description of test development focuses on ISAT, but the ISBE Assessment Division follows the same steps in all the tests it develops.

Relationship of the Illinois Learning Standards to the state assessment system

The Illinois Learning Standards include the written expectations that Illinois students are expected to satisfy in the subjects of language arts, mathematics, science, social science, physical development and health, and fine arts. The first step in constructing tests based on the Standards is to identify the components of each Standard that can be tested in a statewide assessment. For example, Standard 14C, “Understand election processes and responsibilities of citizens,” can be assessed by a multiple-choice question. Many of the Standards overlap and are necessarily connected. Thus, for assessment individual Standards are clustered into sets in each subject and questions are developed to assess these sets.

Developers of the state tests

ISBE’s Assessment Division oversees development and implementation of ISAT, IMAGE, and the PSAE. The task requires the collaborative efforts of Illinois educators and three major contractors.

ISBE Assessment Division Staff

ISBE has a relatively small in-house staff of assessment consultants. Five consultants are responsible for development of the individual tests for each subject assessed by ISAT and the PSAE, while one oversees the development and implementation of the IMAGE reading and writing assessments. Two staff psychometricians provide psychometric design and analysis for all tests. A
psychometrician is a practitioner of the branch of psychology that deals with the design, administration, and interpretation of quantitative tests.

Two staff members provide operational liaison among ISBE consultants and test contractors, oversight of data checking, and coordination of special projects, including Special Education issues, NAEP and TIMSS, and writing and editing services.

**Illinois Educators**

Classroom teachers, curriculum directors, administrators, editors, and content and education specialists are called upon throughout the year for writing test items, selecting test items to be used, and reviewing items, reading passages, writing prompts, and completed tests.

**Major Contractors**

Three major contractors provide the large-scale development, production, and scoring services necessary to the assessment program: **Test development**, which includes the writing of new items and construction of tests, is currently carried out by MetriTech, Inc., of Champaign, Illinois. MetriTech is a major partner of Riverside, for whom it currently scores three test programs. It also does test development for Harcourt and for several states, including Wisconsin, Minnesota, and Oregon. It develops or scores tests for several city school districts, including those in Chicago, Cleveland, Dayton, Cincinnati, Toledo, and Boston.

**Test production**, which includes printing and shipping tests, scoring multiple-choice items, and processing and distributing test data, is currently carried out by NCS Pearson (formerly National Computer Systems) of Iowa City, Iowa. NCS Pearson is the largest commercial processor of student assessment tests for K-12 schools in the U.S. In 2001 the company will score more than 37 million tests. It provides large-scale project management and operations and logistics management for all major U.S. test publishers and for many state departments of education.

The **scoring of writing samples and extended responses items in reading and mathematics** are currently carried out by Measurement, Incorporated, of Durham, North Carolina, which is the largest provider of educational handscoring services in the U.S. It scores essays and extended-response for more than 25 state departments of education. In addition, it has completed other projects for local school districts in many of these states and for other instructional and assessment entities, including the Association of American Medical Colleges, the American College Testing Program, CTB/McGraw-Hill, Harcourt Brace Educational Measurement, the North Carolina Association of Educators, and Riverside Publishing Company.

**Translating Standards to test items**

The Illinois Learning Standards include the written expectations that Illinois students are expected to meet in the subjects of language arts, mathematics, science, social science, physical development and health, and fine arts. (Standards were also adopted for foreign language, but there is no state foreign language assessment.).

The first step in constructing tests based on the Standards was to identify the components of each Standard that could be tested in a statewide assessment. For example, Standard 14C, “Understand election processes and responsibilities of citizens,” can be assessed by multiple-choice, machine-scored questions, but Standard 4B, “Speak effectively using language appropriate to the situation and audience,” is most readily and effectively assessed at the classroom level. Between these two extremes is Standard 3B, “Compose well-organized and coherent writing for specific purposes and audiences,” which can be assessed on a statewide scale but which requires scoring by trained human readers rather than machines.

The next step was to set test specifications. Test construction requires answers to specific questions: How long is the test? How many sessions are there? How should items be grouped? The answers to these questions provide the test specifications.
To construct a test that would adequately assess the standards, it was necessary to group the standards. Many of the standards overlap and are necessarily connected. Thus, standards were clustered into sets in each learning area. For example, there are 18 mathematics standards which include numbers sense, algebra, geometry, statistics, and probability that are grouped into 8 sets, such as algebraic patterns and variables and geometric concepts, that stress the connections within the mathematics content standards and process (problem-solving) standards.

**Cycle of test development**

Optimally, three years are needed to produce each state census test (ISAT, PSAE, IMAGE), which are referred to as census tests because they are given to each student eligible to take them. Items for the test to be given in 2002, must be written in 1999, tried out in 2000, and piloted in 2001. Simultaneously, the preparations for the tests given in other years proceed as well. Thus, in any given year, all phases of test construction are performed. For example, in 2000 the Assessment Division was giving the 2000 census test, writing items for the 2003 census, trying out items for the 2002 census, and administering pilot tests for the 2001 census.

This continual preparation takes places for each grade and subject area for which there is a census test. ISAT includes seven subjects given at five grade levels. At the present time, IMAGE consists of reading and writing tests for three developmental levels with mathematics tests under development. The PSAE covers five subjects.

**Writing Test Items**

The first step in item writing is to decide exactly what type of items are needed. For example, a test may require “items that address algebraic representations at grade 3” or “one new reading passage at grade 5, not involving topics related to animals (as the other passage already does this).” Teachers use item specifications to write actual test questions for the subject and grade that they teach. Item writing is coordinated by MetriTech and guided by ISBE Assessment Division consultants.

After an item is written, it must be reviewed, edited, tried out, and piloted before it can be included on a census test.

**Reviewing Items for Quality**

Review committees, which include teachers at the target grade, screen each item for grade-level appropriateness, content validity, importance, accuracy and possible ambiguity of the answer choices, and clarity of graphics. Teachers review the vocabulary for grade-level appropriateness. Reviewers delete items that are esoteric, trivial, inaccurate, or value-laden.

**Trying Out Items**

Test items and writing prompts are tried out on students in one or two classrooms. Students and teachers then voice their comments in a focus-group setting or comment on the test items on a questionnaire. Responses to the writing prompts are read by scoring staff of Measurement Incorporated. Tryouts eliminate some test items, reading passages, and writing prompts.

**Pilot Testing and Psychometrics**

Items that survive the tryouts and bias review committees are used in a pilot test to determine their psychometric properties.

Specifically, the items are administered to 2,000 or more students, and the results are then analyzed to determine the items’ validity, difficulty, and utility. Also, although the initial review and screening will have removed any items that were believed to be biased or insulting to subgroups of the population, further analyses are performed to ensure that items are also unbiased in a statistical sense. An item is statistically unbiased if students of equal ability (as determined by the number of items they got right on the rest of the test) are equally likely to get that item right.
The validity of an item reflects the degree to which it measures what it is intended to measure. For example, a question about a reading passage is valid if students who read and understood the passage being tested are likely to answer it correctly. The question would not be valid if students who did not read the passage were likely to answer it correctly or if students who did read and understand the passage were likely to answer it incorrectly.

It is necessary to have information about how difficult each item is because items with a range of difficulties are required to construct a test. The utility of an item lies in its ability to distinguish between students who understand the material being tested and students who do not. As students differ considerably with respect to their knowledge of a particular subject area, a wide range of item difficulties is required to reliably assess all students.

Items that survive this statistical screening enter the item bank.

Construction of tests

Test items are selected from a bank of piloted and previously used items. A number of considerations go into selecting items for a test. There must be a range of difficulty in the items so that the test will reflect differences in ability among the students who take it. Generally, items that are too easy (80 percent or more correct responses when the item was piloted) or too difficult (25 percent or fewer correct responses) are not used. The average difficulty of each multiple-choice test is about 60 percent.

Items are selected so that each set of standards is represented by several items. Reading, for example, has six sets of standards, including comprehension of literary works and informational sources, identification of explicit ideas and inferential ideas, and vocabulary questions. Science has five sets of standards, including methods of inquiry, life sciences, physical sciences, earth and space sciences, and science, technology, and society. Mathematics has eight sets of standards, including numbers and computation, algebraic concepts, geometric concepts, measurement, data analysis, and probability. Social science has five sets of standards that include government, economics, geography, U.S. and world history, and global perspectives.

Tests must be comparable from year to year, so new tests include some of the same items that were used in previous years. This permits statistical equating of the data so that scale scores from one year can be compared with those from another.

Test reviewers

Each test item and prompt that is written is reviewed many times before it appears on a census test. After items are assembled into tests, the finished tests are also reviewed. Finally, a committee reviews the results of census tests after they are administered.

Content Advisory Committees

Each subject tested on ISAT, PSAE, and IMAGE is guided by content-specific committees to ensure the validity of the test content, its alignment with the Illinois Learning Standards, and the correlation with innovations and developments in the content area.

Bias Review Committees

Bias review committees review test items, reading passages, and writing prompts before they are used in a test. These committees, which include classroom teachers, administrators, and other educators, examine the items to determine their suitability with respect to cultural bias, community standards, and ethnic sensitivity.

State Testing Review Committee

As required by law, this committee provides feedback from the point of view of the general public. As such, it addresses such issues as, Do the ISAT tests appear inappropriate, irrelevant, or monotonous? Do the questions conflict with cultural or racial sensibilities?
Technical Advisory Committee

The primary responsibilities of this committee are to determine whether

- the various ISAT and IMAGE tests are psychometrically sound;
- the reporting of the test results to students, schools, and district is adequate; and
- the technical reporting of the results is sound.

Logistics of test administration

The final version of the census test is typeset, printed, packaged, and shipped to schools by NCS Pearson. Given the enormous scale of the testing operation, the schedule must be set at least six months in advance so that this work can be completed on schedule.

The ISAT and IMAGE census tests are administered during a two-week period designated by ISBE. Typically, the number of test takers ranges from about 140,000 in grade 3 to 110,000 at grade 11. In 2000, almost 600,000 students were tested.

It is important that all Illinois students sit for the test during the same limited time period so that statewide scores can be compared. It is also important that schools return tests to NCS Pearson by the established deadline. Test materials left in the schools are a risk to test security, and scoring can be seriously delayed if tests are not available as scheduled.

Scoring the state tests

After schools ship the tests back, the multiple-choice and extended-response items are mechanically separated by a slitting machine. The multiple-choice answer sheets are machine-scored by NCS. Essays and extended-response items are sent to Measurement Incorporated for individual scoring by trained readers.

Rubric for Extended-Response Items

A scoring rubric, or prescription for scoring, is used to score the extended-response questions. The rubric for each grade level and subject currently being tested was developed by Illinois teachers working with staff from ISBE and Measurement Incorporated.

Validation Committees and Extended-Response Scoring

Each area that uses an extended-response format has a committee to select and score the papers that are used to train the readers that score the answers. Separate committees exist for reading, writing, and mathematics.

Validation committees constituted of teachers who teach at the grade level being validated are convened to select and score individual papers, referred to as anchor papers because they are used as exemplars for scoring all other test papers. The selected papers are representative of each possible score for each extended-response question or writing prompt. Teachers score papers at the grade level that they teach. Once a sufficient number of anchor papers are selected and scored, Measurement Incorporated begins to train its readers so that their scoring of papers is consistent with the anchor-paper scores. In 1999, a total of 27 different extended-response items were included in ISAT and IMAGE tests.

Measurement Incorporated hires about 1,300 staff members to read and score approximately 3 million responses, a process that takes at least 6 weeks. Measurement Incorporated begins the scoring process by recruiting scorers who are knowledgeable in the subject. Many are retired teachers, and many have previous experience with extended-response scoring. All have at least a bachelor's degree and about a third have advanced degrees.

The scorers are trained on the rubric and anchor papers, then required to demonstrate accurate scoring before they begin scoring student papers. Scorers who qualify are assigned to a team with a team leader who supervises the procedure and is responsible for the reliability of the scores. Team
leaders read papers flagged by scorers for help and spot-check scores to make sure that scorers continue to be accurate. There are other checks on the scorers: about 10 percent of papers are scored by more than one scorer, and anchor papers are circulated with unscored student work to be rescored. When significant discrepancies are detected, scorers are retrained and papers rescored. This procedure results in scores with a rater agreement in excess of 0.85.

Data checking: reporting of accurate results

Before scores are reported back to schools, NCS Pearson must process the multiple-choice scores, and Measurement Incorporated must score the extended-response items and send those scores to NCS Pearson. NCS Pearson then sorts and assembles the data into school and district scores and individual student reports. If all goes well, the necessary steps from testing to reporting require about 3½ months. This length of time is necessary because of the complexity of the state tests and to ensure accuracy of reported results. ISBE staff and the ISBE contractors check the correctness of the ISAT results in a number of ways.

- All tests are reviewed several times by ISBE staff before being printed to ensure that scoring keys are correct.
- The correctness of the scoring keys is rechecked once the tests are returned to the scoring contractor by reviewing the answers of the highest and lowest scoring students.
- Before any ISAT reports are sent out to districts, schools, or students, ISBE staff checks the data of a small number of school and districts. This involves doing by hand the computations that the computer performs, a process that takes two to three weeks.
- Finally, schools and districts are asked to verify the preliminary data and key pieces of information. Typical problems uncovered are miscoding students by category (ethnic, gender, disability, and other classifications), erroneously assigned grade levels, and missing data.

ISBE and the printing contractor correct any erroneous information that schools make known to them. Finally, ISBE checks each school’s rate of participation in ISAT by comparing the number tested to the number the school reports as enrolled at the time of testing.

Performance levels for ISAT and how they are set

The Illinois Learning Standards describe what students are expected to master at five benchmark levels (early elementary, late elementary, middle and junior high school, early high school and late high school levels) in each of the seven areas addressed by the ISATs and PSAE. Student performance is categorized either as Exceeds Standards, Meets Standards, Below Standards, or as sufficiently poor to be classified as being at Academic Warning. The distinction between meeting Standards (i.e., the Meets + Exceeds levels) and not meeting Standards (i.e., Below Standards + Academic Warning levels) is the most crucial.

Once the Illinois Learning Standards were available, they were applied for ISAT using a two-step process. First, a committee of subject-area experts, including teachers working at the grade level that they were evaluating, convened to write performance definitions that describe the four performance levels unambiguously in terms of the skills and knowledge needed to attain them. Although such statements clearly delineate what is expected, the definitions still must be set in terms of actual ISAT score ranges, or cut scores, for each area and grade.

To accomplish this step, a second committee of teachers was convened for several days to set each cut score. The standard-setting committees comprised a cross section of teachers working at the grade level under discussion (one group per grade and area) that reflected the geographical and ethnic differences in the state.

State Assessment System: DRAFT VERSION
To set the ISAT cut scores, a modified-Angoff method was used. This method involves the careful review of each item in a test to estimate how well students at each of the four performance levels would perform. Committee members are instructed to imagine students whom they have taught and who match the performance level as described in the performance definition and given the opportunity to discuss students in this range to firm up their conceptions. Then they are asked to estimate for each question the percentage of students at that grade level who would be expected to answer correctly.

To assist the members of the ISAT standard-setting committee in their task, committee members were given information about students’ actual performance on each item (the percent of students getting the correct answer). Next, the committee members focused on the differences among students who exceed, meet, or don’t meet Standards, or who are classified as in academic warning. For example, to determine the ISAT score that separates meeting from not meeting the standards, committees focused on 100 hypothetical students whose performance they considered to be at the border between these categories (i.e., Academic Warning vs. Below, Below vs. Meet, and Meet vs. Exceeds). Each ISAT item was then discussed extensively to obtain informed expert estimates. These estimates were then translated into ISAT scores by statistical means.

Teachers showed high agreement in their estimates. By aggregating the committees’ estimates across all test items, ISAT cut scores were computed to define each standard level. As year-to-year ISAT test scores are equated over, the cut scores remain fixed and it is not necessary to repeat the standard setting procedure on a yearly basis. The cut scores for each grade and area follow:

**READING (Scale= 120-200)**

<table>
<thead>
<tr>
<th>GRADE</th>
<th>Academic Warning</th>
<th>Below Standards</th>
<th>Meets Standards</th>
<th>Exceeds Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>120-137</td>
<td>138-155</td>
<td>156-173</td>
<td>174-200</td>
</tr>
<tr>
<td>5</td>
<td>120-129</td>
<td>130-155</td>
<td>156-170</td>
<td>171-200</td>
</tr>
<tr>
<td>8</td>
<td>120-128</td>
<td>129-151</td>
<td>152-172</td>
<td>173-200</td>
</tr>
<tr>
<td>10</td>
<td>120-135</td>
<td>136-152</td>
<td>153-174</td>
<td>175-200</td>
</tr>
</tbody>
</table>

**WRITING (Scale= 6-32)**

<table>
<thead>
<tr>
<th>GRADE</th>
<th>Academic Warning</th>
<th>Below Standards</th>
<th>Meets Standards</th>
<th>Exceeds Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6-13</td>
<td>14-21</td>
<td>22-29</td>
<td>30-32</td>
</tr>
<tr>
<td>5</td>
<td>6-13</td>
<td>14-20</td>
<td>21-27</td>
<td>28-32</td>
</tr>
<tr>
<td>8</td>
<td>6-14</td>
<td>15-20</td>
<td>21-27</td>
<td>28-32</td>
</tr>
<tr>
<td>10</td>
<td>6-14</td>
<td>15-20</td>
<td>21-27</td>
<td>28-32</td>
</tr>
</tbody>
</table>

**MATHEMATICS (Scale= 120-200)**

<table>
<thead>
<tr>
<th>GRADE</th>
<th>Academic Warning</th>
<th>Below Standards</th>
<th>Meets Standards</th>
<th>Exceeds Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>120-141</td>
<td>142-152</td>
<td>153-172</td>
<td>173-200</td>
</tr>
<tr>
<td>5</td>
<td>120-137</td>
<td>138-157</td>
<td>158-190</td>
<td>191-200</td>
</tr>
<tr>
<td>8</td>
<td>120-137</td>
<td>138-161</td>
<td>162-184</td>
<td>185-200</td>
</tr>
<tr>
<td>10</td>
<td>120-138</td>
<td>139-157</td>
<td>158-187</td>
<td>188-200</td>
</tr>
</tbody>
</table>

State Assessment System: DRAFT VERSION
Other Assessment Issues

As the stakes associated with testing have risen, stakeholders — including educators, parents, legislators, and members of the public — have taken a greater interest in assessment issues. There are debates about the optimum time for testing, the need for standardized conditions and higher levels of security, the fairness of holding schools accountable for conditions they cannot completely control, the quality of ISBE-developed tests vs. off-the-shelf standardized tests, and many other issues. This section addresses some of these.

Test security

ISBE proceeds with considerable confidence in the trustworthiness of Illinois teachers and administrators, an approach that appears to be warranted. ISBE receives only 10 to 20 allegations of irregularities a year only a few of which are allegations of cheating for over 40,000 testing rooms and 850,000 students tested. Follow-up is done for all allegations, and there have been fewer than 10 substantiated allegations of cheating during the 13 years of state testing from 1988 through 2000.

There is a system of checks on the integrity of ISAT administration. In cases of highly questionable similarities among essays or extended responses, ISBE regularly asks school administrators to investigate the matter and attaches a copy of ISBE’s rules for investigating allegations of irregular incidents to the request. The expectation is that staff responsible for the irregularity will be appropriately disciplined. If irregularities are substantiated, ISBE will suppress ISAT scores for that year for the school or district affected. Ultimately, this will damage their designation status. Moreover, the contents of ISAT and the PSAE are changed after each test administration, in contrast to the standardized norm-referenced tests developed for national dissemination, which usually change only after several years of administration. The test questions on those tests remain identical and are thus far easier to remember from year to year.

Test irregularities may affect a school’s designation: The proposed school designation model will establish criteria for inclusion of schools and districts on the Academic Early Warning List. Part of the compliance prerequisites for this new model will include sanctions for cheating or other inappropriate practices. For this reason, the ISBE Assessment Division has prepared and distributed a document, Professional Testing Practices for Educators, to ensure that administrators and teachers are aware of testing practices that — if followed — will result in standardized testing conditions. The eight-page document summarizes required, prohibited, and strongly recommended practices.

Educators who have become familiar with the testing practices that are described in this document will be able to ensure that their students take the ISAT under fair and equitable conditions. To achieve this goal, Illinois public school educators are expected to read the testing practices document carefully and to adhere strictly to the principles and policies it describes. Superintendents and principles are required to sign a test security and confidentiality document to indicate that they have taken sufficient and appropriate steps to ensure that all who handle secure test materials in the school are fully aware of their responsibilities regarding test security and confidentiality and are properly trained to handle these responsibilities.

Security procedures for the PSAE are extremely rigorous. The test is given using ACT’s stringent test-administration protocol. Thus, this higher-stakes test also has higher security. It would difficult to administer ISAT under conditions as strict, given the demand for more flexibility with numerous grades and subjects.

Illinois Learning Standards in relationship to the time of testing

While Illinois does not have a state curriculum, it has Learning Standards that were adopted by the State Board in July 1997, after considerable input and development by thousands of Illinois citizens. These standards are widely publicized and can be found at the ISBE Web site. The Learning Standards
are not intended to define end-of-year exams. Rather, they reflect what students need to master cumulatively at certain points in their school careers at the time the state tests are given. Teachers developed the questions based on what they think students should know by the testing date. If some schools do not get around to teaching some of the material included on the test, they need to make changes in their curriculum to accelerate coverage of the material.

The PSAE test questions are based upon learning standards that should be included in a variety of courses offered to students prior to grade 11. To illustrate, a student does not have to wait until enrollment in a grade 12 elective economics course to learn knowledge required to answer a question based upon the Learning Standard “understand trade as an exchange of goods and services.” This is an economic concept that should have been taught in history and geography classes many times throughout a student’s academic career before grade 12. As pointed out earlier, the quality of all ISAT and PSAE questions are verified via pilot tests in many schools across the state. The fact that the questions are written by teachers ensures that they are appropriate to the grade levels at which they are used.

Getting scores back: How schools use scores

Schools have two irreconcilable desires: (1) to test students as late as possible in the school year and (2) to have test results before the end of the school year. The turn-around time for test results is a major problem in statewide testing across the country. Many people demand a quick return of results in the mistaken belief that all tests are the same and that data can be returned equally quickly from any standardized test.

Several factors contribute to the length of the interval between testing and reporting for Illinois state assessments:

- Enormous scale of the testing – more than 700,000 students are tested each year.
- Complexity of the tests, which include extended responses in reading and mathematics as well as writing samples. After NCS Pearson receives completed answer documents, it must separate extended-response pages from multiple-choice pages and send each part to the proper location for scoring.
- Scoring of extended-response items and writing require more than six weeks and take place at multiple centers throughout the country.
- Delay in return of completed tests by some schools. With more than 3.5 million test booklets to score, NCS Pearson cannot complete even the machine scoring in a timely manner unless schools ship the completed tests on the day testing is completed. After scoring is completed, NCS Pearson must match the extended-response scores with the multiple-choice scores for each student.
- Numerous procedures are in place to check the accuracy of the generated data.
- Finally, test scores must be equated each year so that they are comparable from year to year.

ISAT is only one part of the testing system that schools use for planning their curricula. It is not designed primarily as a diagnostic test for individual students. In fact, a single test cannot be and should not be used for diagnostic purposes. True diagnosis, whether it is in education or the medical field, requires many types of assessment to determine what may be needed to improve academic or physical health, respectively. ISAT provides accountability at the school, district, and state levels, and its purpose is to guide the program-development efforts of schools and districts. ISAT provides reliable data that can be compared from year to year, which schools can use to assess the effectiveness of curricular modifications.
An extended quote from an article by Stuart Kahl published on the National Association of State Boards of Education Web site in which he states that many states encounter the call of faster test results (http://www.nasbe.org/stakes_mistakes.html) describes the test-return problem well:

A common concern expressed by many local school personnel regarding statewide testing programs is the time it takes for them to get test results. Their frame of reference is their experience in administering off-the-shelf tests as part of school or district testing programs. These tests are developed over a period of years, studies are completed to determine national norms, and scoring and reporting programs are perfected before the tests are put on the market. Schools administer the tests, then return their students answer sheets, and within weeks receive reports. Because of the advance work that has already been completed, computers do little more than determine points earned by students, then look up in a table the corresponding scaled scores by going to the right place for the age of the students and time of year tested. This is much like using a tax table once adjusted gross income is known.

In the scenario above, schools receive student reports as well as school summaries within a relatively short time period. For local testing, it does not matter whether the test published receives answer documents for all students within a school. The publisher scans and scores whatever materials it was sent and returns results for those students who materials it processed; having no knowledge of whether students were inappropriately excluded or otherwise unaccounted for. However, the situation is usually different for statewide testing. In many state programs, the contractor spends a great deal of time tracking down information on students when the numbers of student testing materials do not agree with the enrollment figures of the schools. More often than not, this happens whether the statewide program uses customized or off-the-shelf instruments. Since many programs produce current-year statewide results for purposes and comparisons in the school reports, all school and student materials must be accounted for before final results for the state can be determined. The goal has two inherently contradictory requirements: complete information on all districts, schools, and students; and fast, efficient reporting of results. Many tasks must be accomplished before reports of results can be distributed. If a program uses mixed item formats, then multiple-choice responses can be processed immediately, but answers to constructed-response questions must be scored before data from those questions can be merged with data from the multiple-choice component. The merging of files generally uncovers the need for additional data cleanup because of unmatched partial records. If the program is totally customized assessment program, using new tests every year, existing computer programs must be modified, and sophisticated scaling and equating procedures must be implemented each year. Then after all reportable statistics have been computed, report programs must be run to transfer the statistics from the data file to the report shells or formats. With some testing programs producing many different reports of student, school, and district results, the pages of unique data generated may number in the millions. Smudging and other printing problems may create the need for time-consuming return.

Mr. Kahl goes on to say that there is a major irony in the demand for the quick turn-around time. Because the primary purpose of state assessment is to monitor change at the state and school level, several years of data should be grouped to provide good baseline information and the changes reviewed over several years. Results of true reform and improvement efforts will not be seen in one year, but rather will occur as a trend a number of years after the reforms are implemented.

**Technical quality of the ISAT and PSAE**

The products of the state testing system are tests that are equal to or better than commercially produced, or off-the-shelf, tests with respect to cost and technical quality. For instance, the following table reports the reliability of the various 1999 ISAT tests by grade.
Table 1: ISAT Reliability

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.95</td>
<td>0.94</td>
<td>0.95</td>
<td>0.93</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>5</td>
<td>0.93</td>
<td>0.91</td>
<td>0.95</td>
<td>0.94</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.91</td>
<td>0.93</td>
</tr>
<tr>
<td>8</td>
<td>0.93</td>
<td>0.91</td>
<td>0.95</td>
<td>0.95</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows that the reliability of the ISATs ranges from 0.87 to 0.95, values are equal to those of commercially available tests. For instance, the reliability of the California Achievement Test ranges from 0.84 to 0.96 for the reading subtests and from 0.84 to 0.96 for the mathematics score.

There is also solid evidence for the validity of the ISATs relative to commercial tests. The correlation between students’ scores on the reading and mathematics ISAT with their scores on the Stanford Achievement Test (SAT9) are shown in Table 2. (Note: The SAT9 is administered to a sample of Illinois students to determine the percent of students in each national SAT9 performance quarter).

Table 2: Correlation between ISAT and SAT9

<table>
<thead>
<tr>
<th>Grade</th>
<th>Reading</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td>5</td>
<td>0.81</td>
<td>0.87</td>
</tr>
<tr>
<td>8</td>
<td>0.80</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Table 2 shows that correlations exceed 0.83, indicating that ISAT and SAT9 scores measure highly similar concepts in reading and mathematics. However, the ISAT is directly linked to the Illinois Learning Standards, so it is preferred for state accountability purposes.

Reliability of scores

The General Assembly passed legislation that requires all Illinois students to take the ISAT and PSAE, so no student sampling is performed, and thus by definition, no sampling errors can occur. As a result, the often-cited rule for sample sizes (i.e., the size of the sample must be 30 or larger) does not apply: ISAT concerns schools’ populations, not samples of those populations. Thus, differences between schools over time are not the result of sampling error, but rather reflect genuine differences in performance of the students in a particular year.

From the schools’ perspectives, scores are sometimes higher and sometimes lower than expected. Regardless of the size of the schools, however, these variations are real. We know that schools sometimes seem to have a class of students that perform either much better or much worse than the classes before or after them. This effect can be particularly noticeable in small schools. To strike a balance between unusually good and poor classes, ISBE watch and warning list designations have been based on multiple years of data, and the school designations are based on the data for all subjects across all grades.
Regardless of the size of the school, all students must be held to the same standard. Students graduating from high school and competing for jobs or places in colleges and universities will not be given special consideration based on the size or location of their schools. It would be highly discriminatory to require some schools to have higher standards for students than other schools.

**Mobility of students**

Illinois law requires that all students be tested in grades 3, 4, 5, 7, 8, and 11, including secondary language learners and poor and highly mobile students. The correlation between student mobility and poverty is 0.64, which means that the majority of schools with high mobility rates are also low income. Many of these students do not leave a school on a permanent basis, but rather move from one school to another within the same district and then back as their families’ monthly apartment rents fall due. Recognizing that schools have a much more difficult time with this type of chronic mobility than with students who move in or out of a school once in a school year, the new system will take mobility into account when assigning school designations.

Because of concerns about student mobility, ISBE initiated a study that will compare the 2001 ISAT scores of students who enter school at the beginning of the school year with those who enter after October 1 and after January 1. A finding of systematically low scores for high mobility students would indicate that greater effort is needed to bring them to the level established by the Illinois Learning Standards. Schools and districts with such subgroups can legitimately point out these facts to their constituents when explaining their ISAT results. Currently, there is legislation in the General Assembly to not include scores of students who enter a school after October 1.

The recently activated ILSI Web site at [http://ilsi.isbe.net](http://ilsi.isbe.net) provides extensive information for all Illinois schools with respect to student mobility. In particular, it provides schools with a means to compare themselves to other schools with similar levels of student mobility.

**Opportunities for makeup testing and inclusion of students with disabilities and limited English proficiency**

There are two weeks during which ISAT can be given, and makeup tests are included in that time period. Each Illinois school receives a testing manual, which explains that there is a 10-day window for administering ISAT. This was done to allow schools to do their main test administration early and then reschedule makeups during the remaining days. The new designation system will require a 90-percent-tested rate in order to be a Meets school; scores of schools not reaching the 90 percent level will be flagged, and the schools investigated. Currently, the percent-tested report does not include those students who take IMAGE or the IAA. Some schools with large numbers of students taking IMAGE may have a lower percent tested owing to the ISAT-only reporting in the current Report Card.

The decisions as to which students with disabilities who have Individualized Education Programs (IEPs) should or should not take ISAT are to be made on a case-by-case basis. Regardless of an IEP student’s disability category (for example, deaf, blind, LD), local educators should make the decision about the individual student’s participation in ISAT (and any necessary accommodations) and document their decision in the student’s IEP. As mandated by the Individuals with Disabilities Education Act (IDEA) Amendments of 1997, these decisions must never be driven by any sort of blanket directive or one-size-fits-all approach. Decisions should never be based on the category of a student’s disability, the setting in which the student receives instruction, the percent of time the student spends in a particular classroom, or expectations that the student will not perform well on a particular assessment. Decisions should instead be based on the learning characteristics of the individual student and the task demands of a given assessment. Students who do not participate in ISAT will take the Illinois Alternate Assessment.

Scores of students with IEPs are to be reported to their home schools. If there is a concentration of LD students in a building, it is likely because that school houses a program that receives students from
other schools. Thus, the scores for those students would not be reported to the school where they attend class but to their home schools.

To provide a bridge to the ISAT, newly arrived foreign language students may take a special test (IMAGE) for up to three years until they become proficient in English. IMAGE assesses students’ ability to read and write in English. Clear criteria have been established to guide students’ transition from IMAGE to ISAT, and this will streamline the integration of bilingual students into the Illinois school system. The inclusion of mathematics in IMAGE is currently under study. IMAGE scores are **not** included in the ISAT results but are linked to the Illinois Learning Standards.

The inclusion of all students is mandated by law for all categories of students. Starting with the 2001 ISAT, school and district reports will provide more detailed information about the percentage of students tested by category (Title 1, LEP, IEP, gender, ethnicity). This detailed reporting of participation is mandated by federal Title 1 regulations.

Schools’ percentages of students tested are currently being considered as a criterion in the new school-designation system.
ISAT, PSAE, and IMAGE – A Fact Sheet

- The development of each new census test takes three years.
- More than 3.5 million test booklets are needed for all three tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>Different Documents</th>
<th>Different Pages</th>
<th>Field Tests</th>
<th>Documents Printed</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISAT</td>
<td>34</td>
<td>2200</td>
<td>80</td>
<td>3,000,000</td>
</tr>
<tr>
<td>PSAE</td>
<td>74</td>
<td>930</td>
<td>74</td>
<td>340,000</td>
</tr>
<tr>
<td>IMAGE</td>
<td>6</td>
<td>700</td>
<td>45</td>
<td>163,000</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>3830</td>
<td>199</td>
<td>3,503,000</td>
</tr>
</tbody>
</table>

Data provided by NCS Pearson for 2000 census test

- About 20 semitrailers are needed to transport the test materials from NCS in Iowa City to Illinois.
- These test booklets are distributed to about 3,500 Illinois schools for use by some 700,000 students.
- ISAT reports are sent to the parents or guardians of more than 500,000 Illinois students.
- Measurement Incorporated scores more than 3 million extended-response questions in about 6 weeks.
- MetriTech, Inc., analyzes the psychometric properties of more than 200 test forms.