The Illinois State Assessment
2005 Technical Manual

Illinois State Board of Education
Division of Assessment

CONTENTS

1. PURPOSE AND DESIGN OF THE ISAT TESTING PROGRAM

In April 2005, students in grades 3, 5, and 8 took Illinois Standards Achievement Tests (ISAT) in reading and mathematics. Students in grades 4 and 7 took ISAT tests in science. Approximately 750,000 students enrolled in public elementary and secondary schools across the state participated in the testing program. ISAT measures the extent to which students are meeting the Illinois Learning Standards. Illinois teachers and curriculum experts developed the ISAT tests in cooperation with the Illinois State Board of Education (ISBE).

This manual provides technical information about the 2005 test administration. It describes the tests and assessment approaches and addresses technical concerns. Other reports, documents, or publications issued by the Illinois State Board of Education (ISBE) provide additional information about interpreting test results (*Guide to the 2005 Illinois State Assessment*, *Understanding Your Child's ISAT Scores*) that is not included here.

General Procedures

Each ISAT test is designed to ensure that its results validly and fairly assess the Illinois Learning Standards. The selection of items and assembly of each test is guided by a set of specifications. These specifications were developed by Illinois educators to help ensure that test content corresponds to the purposes, objectives, and skills framed by the learning standards.

Illinois teachers and administrators participate in all phases of the test development process: item writing, item selection, bias review, and test assembly. The State Board of Education 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 Illinois program:

> *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. In tests other than reading, 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 prior to their inclusion in tests to develop a statistical profile for each item. 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 .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.

ISBE takes several precautions to help ensure test security. Test materials shipped to schools are packaged and sealed. Each test booklet is bar-coded so that it can be accounted for. The administration of tests is standardized. A series of manuals provides guidance on security and other issues to the district testing coordinator, school testing coordinator, and classroom test administrator. After administration, all materials are removed from schools and returned to a central facility for processing and secure destruction of unneeded materials.

Reading

The ISAT reading test assesses material defined by standards associated with three state learning goals. The standards were developed using the 1985 State Goals for Language Arts, various state and national standards drafts, and local education standards contributed by team members. These learning standards are designed to guide language arts instruction in Illinois schools. This alignment of assessment to curriculum ensures consistency and strengthens the influence of standards and assessment on improved teaching and learning. These standards are:

- Goal 1: Read with understanding and fluency.
  - 1A Apply word analysis and vocabulary skills to comprehend selections.
  - 1B Apply reading strategies to improve understanding and fluency.
  - 1C Comprehend a broad range of reading materials.

- Goal 2: Read and understand literature representative of various societies, eras, and ideas.
  - 2A Understand how literary elements and techniques are used to convey meaning.
  - 2B Read and interpret a variety of literary works.

- Goal 5: Write to communicate for a variety of purposes.
  - 5A Locate, organize, and use information from various sources to answer questions, solve problems, and communicate ideas.
  - 5B Analyze and evaluate information acquired from various sources.
  - 5C Apply acquired information, concepts, and ideas to communicate in a variety of formats.

The reading test has two formats. The grade 3 reading assessment is given in three 40-minute sessions. One of these sessions consists of 12-15 word analysis questions and one passage followed by 15-17 multiple-choice questions. The two remaining sessions include one passage followed by 15-20 multiple-choice questions and one extended-response question.

The reading tests for grades 5 and 8 are also given in three 40-minute sessions. One of these sessions consists of a longer passage with 20-25 multiple-choice questions. The other two sessions each include one passage with 15-20 multiple choice questions and one extended-response question.

The reading passages and accompanying questions reflect two of the most frequent purposes for reading—reading to gain information and reading for literary experience. The sources for these passages range from high interest, grade-appropriate periodicals to newspapers, short stories, and novels. Illinois teachers reviewed and selected the material for these tests.

The multiple-choice questions require students to select one correct response from four possibilities presented to them. Again, teachers in Illinois played an active part in writing, reading, and editing these test questions. Questions must meet both content and statistical criteria for inclusion in the test.

The extended-response questions on the reading test require students not only to read and understand a text, but also to analyze, evaluate, and interpret the text as a means of making connections and conclusions related to the text. The rubric used to score the extended-response items is a holistic scoring rubric. It describes characteristics of different levels of achievement in reading. The levels of achievement on the reading rubric range from 0 to 4 (4 being the highest score). Responses with scores of 0 indicate that the student response is insufficient to effectively determine evidence of achievement in reading. Responses with scores of 1 and 2 indicate developing levels of achievement in reading. Responses with scores of 3 indicate a developed level of achievement in reading. Finally, responses with scores of 4 represent a well-developed level of achievement in reading. The rubric was developed with Illinois educators.

In addition to an overall reading score, results are reported in terms of the percent of items correctly answered within five "standard sets" (six at grade 3). These scores are as follows:

- *Comprehension: Literary Works:* Understanding of passages taken from sources such as novels, short stories, and periodicals. (Standards 1B, 1C, 2A, 2B, 5A, 5B, 5C)

- *Comprehension: Informational Sources:* Understanding of nonfiction texts such as student periodicals, newspapers, and trade journals. (Standards 1B, 1C, 2A, 2B, 5A, 5B, 5C)

- *Application of Strategies: Explicit Ideas:* Identifying important information directly stated in the text. (Standards 1B, 5A)

- *Application of Strategies: Inferences from Text:* Analyzing important information in the text to draw logical conclusions about the text. (Standards 1C, 2A, 2B, 5B, 5C)

- *Vocabulary:* Using contextual clues and other skills to understand key words, phrases, and concepts in literary and informational texts. (Standard 1A)

- *Word Analysis (3rd grade only):* Using phonics, word pattern, and other word analysis skills to recognize new words. (Standard 1A)

Mathematics

People use mathematics to identify, describe, and investigate the patterns and challenges of everyday living. Mathematics helps us to understand events that have occurred and to predict and prepare for events to come so that we can more fully understand our world and more successfully live in it. Mathematics encompasses arithmetic, measurement, algebra, geometry, trigonometry, statistics, probability, and other fields. It deals with numbers, quantities, shapes, and data, as well as numerical relationships and operations. Confronting, understanding, and solving problems are at the heart of mathematics. Mathematics is much more than a collection of concepts and skills; it is a way of approaching new challenges through investigating, reasoning, visualizing, and problem-solving with the goal of communicating the relationships observed and problems solved to others.

The ISAT mathematics tests are designed to measure the following learning standards:

- Goal 6: Demonstrate and apply a knowledge and sense of numbers, including numeration and operations (addition, subtraction, multiplication, division), patterns, ratios, and proportions.
  - 6A Demonstrate knowledge and use of numbers and their representations in a broad range of theoretical and practical settings.
  - 6B Investigate, represent, and solve problems using number facts, operations (addition, subtraction, multiplication, division) and their properties, algorithms, and relationships.
  - 6C Compute and estimate using mental mathematics, paper-and-pencil methods, calculators, and computers.
  - 6D Solve problems using comparison of quantities, ratios, proportions, and percents.

- Goal 7: Estimate, make, and use measurements of objects, quantities, and relationships and determine acceptable levels of accuracy.
  - 7A Measure and compare quantities using appropriate units, instruments, and methods.
  - 7B Estimate measurements and determine acceptable levels of accuracy.
  - 7C Select and use appropriate technology, instruments, and formulas to solve problems, interpret results, and communicate findings.

- Goal 8: Use algebraic and analytical methods to identify and describe patterns and relationships in data, solve problems, and predict results.
  - 8A Describe numerical relationships using variables and patterns.
  - 8B Interpret and describe numerical relationships using tables, graphs, and symbols.
  - 8C Solve problems using systems of numbers and their properties.
  - 8D Use algebraic concepts and procedures to represent and solve problems.

- Goal 9: Use geometric methods to analyze, categorize, and draw conclusions about points, lines, planes, and space.
  - 9A Demonstrate and apply geometric concepts involving points, lines, planes, and space.
  - 9B Identify, describe, classify, and compare relationships using points, lines, planes, and solids.
  - 9C Construct convincing arguments and proofs to solve problems.
  - 9D Use trigonometric ratios and circular functions to solve problems.

- Goal 10: Collect, organize, and analyze data using statistical methods; predict results; and interpret uncertainty using concepts of probability.
  - 10A Organize, describe, and make predictions from existing data.
  - 10B Formulate questions, design data collection methods, gather and analyze data, and communicate findings.
  - 10C Determine, describe, and apply the probabilities of events.

Illinois teachers developed the Illinois Learning Standards for mathematics. These goals, standards, and benchmarks are an outgrowth of the 1985 Illinois State Goals for Learning influenced by the latest thinking in school mathematics. This includes the National Council of Teachers of Mathematics; Curriculum and Evaluation Standards for School Mathematics; ideas underlying recent local and national curriculum projects; results of state, national, and international assessment findings; and the work and experiences of Illinois school districts and teachers.

The mathematics assessment includes 70 scored multiple-choice items administered in two test sessions. A third session contains two extended-response/problem-solving tasks.

In addition to an overall mathematics score, results are reported in terms of the percent of items correctly answered within eight standard sets. These scores are as follows:

- *Estimation/Number Sense/Computation:* Demonstrating an understanding of numbers, their representations, and number operations of addition, subtraction, multiplication, division, percentages, and fractions as appropriate to grade level. (Standards 6A, 6B, 6C, 6D, 8C)

- *Algebraic Patterns/Variables:* Identifying, describing, and extending algebraic, geometric, and numeric patterns and constructing and solving problems using variables. (Standards 8A, 8D)

- *Algebraic Relationships/Representations:* Representing and interpreting algebraic concepts with words, diagrams, tables, coordinate graphs, equations, and inequalities. (Standard 8B)

- *Geometric Concepts:* Identifying and describing points, lines, two- and three-dimensional shapes and their properties, such as parallel; symmetry; perpendicular; and number of sides, faces, and vertices. (Standard 9A)

- *Geometric Relationships:* Sorting, classifying, comparing, and contrasting geometric figures. This category includes such properties as similarity and congruency. (Standards 9B, 9D)

- *Measurement:* Estimating, measuring, and comparing quantities using appropriate units and acceptable levels of accuracy. At higher grades, this category encompasses conversions within measurement systems. (Standards 7A, 7B, 7C)

- *Data Organization/Analysis:* Creating, analyzing, displaying, and interpreting data using a variety of graphs (pictures, tallies, tables, charts, bar graphs, Venn diagrams), and computing the mean, median, mode, and range of given data. (Standards 10A, 10B)

- *Probability:* Determining, describing, and applying elementary probability theory and fundamental counting principles. At higher grades, this category encompasses combinations and permutations of simple and complex events. (Standard 10C)

Science

Science is a creative endeavor of the human mind. It offers a special perspective on the natural world in terms of understanding and interaction. The Illinois Learning Standards for science are organized by goals that inform one another and depend upon one another for meaning. Expectations for learners related to the inquiry process are presented in standards addressing the application of science and elements of technological design.

The ISAT science tests are designed to measure the following three learning standards.

- Goal 11: Understand the process of scientific inquiry and technological design to investigate questions, conduct experiments, and solve problems.
  - 11A Know and apply the concepts, principles, and processes of scientific inquiry.
  - 11B Know and apply the concepts, principles, and processes of technological design.

- Goal 12: Understand the fundamental concepts, principles, and interconnections of the life, physical, and earth/space sciences.
  - 12A Know and apply concepts that explain how living things function, adapt, and change.
  - 12B Know and apply concepts that describe how living things interact with each other and with their environment.
  - 12C Know and apply concepts that describe properties of matter and energy and the interactions between them.
  - 12D Know and apply concepts that describe force and motion and the principles that explain them.
  - 12E Know and apply concepts that describe the features and processes of Earth and its resources.
  - 12F Know and apply concepts that explain the composition and structure of the universe and Earth's place in it.

- Goal 13: Understand the relationships among science, technology, and society in historical and contemporary contexts.
  - 13A Know and apply the accepted practices of science.
  - 13B Know and apply concepts that describe the interaction between science, technology, and society.

The science assessment consists of single-correct-answer, multiple-choice items. In addition to an overall score, results are reported in terms of the percent of items correctly answered within five standard sets. These scores are as follows:

- *Scientific Inquiry:* Understanding and applying knowledge of experimental and technological design including data analysis, use of scientific instruments, and the metric system. (Standards 11A and 11B)

- *Life Sciences:* Understanding and applying knowledge of biology and ecology. (Standards 12A and 12B)

- *Physical Sciences:* Understanding and applying knowledge of chemistry and physics. (Standards 12C and 12D)

- *Earth and Space Sciences:* Understanding and applying knowledge of geology, weather, renewable resources, astronomy, and space science. (Standards 12E and 12F)

- *Science, Technology, and Society:* Understanding and applying knowledge of safety, valid sources of data, and ethical practices. Understanding and applying knowledge of the history and sociology of science, ethics, environmental issues, and recycling. (Standards 13A and 13B)

A set of science pilot items and a set of health/physical development items used for conducting state studies bring the total number of items in each test to 80. The pilot items do not contribute to test scores.

The Productive Thinking Scale (PTS) is used to evaluate the quality of science items. It is hierarchical with respect to the production of knowledge and independent of an item's difficulty or grade. Four cognitive skills define the hierarchy of productive thinking in generating scientific knowledge. Each skill applies to both content (knowledge) and to 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. Hence, the PTS 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 which most students must use to answer an item, each item is ranked as to the level of conceptual 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 of items within the tests among major learning areas: earth science, physical science, and life science.

2. RELIABILITY

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 a person's scores are likely to shift unpredictably from one time to another.

Internal Consistency of Overall Scores

Because the items used in achievement tests represent only a relatively small sample from a much larger domain of items, the consistency of test scores across items is of particular interest. That is, how precisely will tests rank students if different sets of items from the same domain were 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*. Table 2.1 presents internal consistency values (coefficient alpha) for each of the tests administered in the assessment.

Table 2.1
Reliability Estimates

Grade	Reading	Mathematics	Science
03	.94	.94	
04			.92
05	.91	.95	
07			.91
08	.92	.96	

Note: Sample sizes on which these coefficients are based are as follows:

Reading: 3 (15,966), 5 (15,955), 8 (15,939)
Mathematics: 3 (15,974), 5 (15,960), 8 (15,946)
Science: 4 (15,971), 7 (15,954)

The reliability coefficients reported in Table 2.1 are 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 four ISAT reading tests; Figures 2.2 and 2.3 present comparable information for the ISAT mathematics tests and science tests, respectively.

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 is evident from the figures, the information functions for these tests are highest near the points on the scales where the "meets standards" cut scores are located.

Figure 2.1
ISAT Reading Test Information Functions

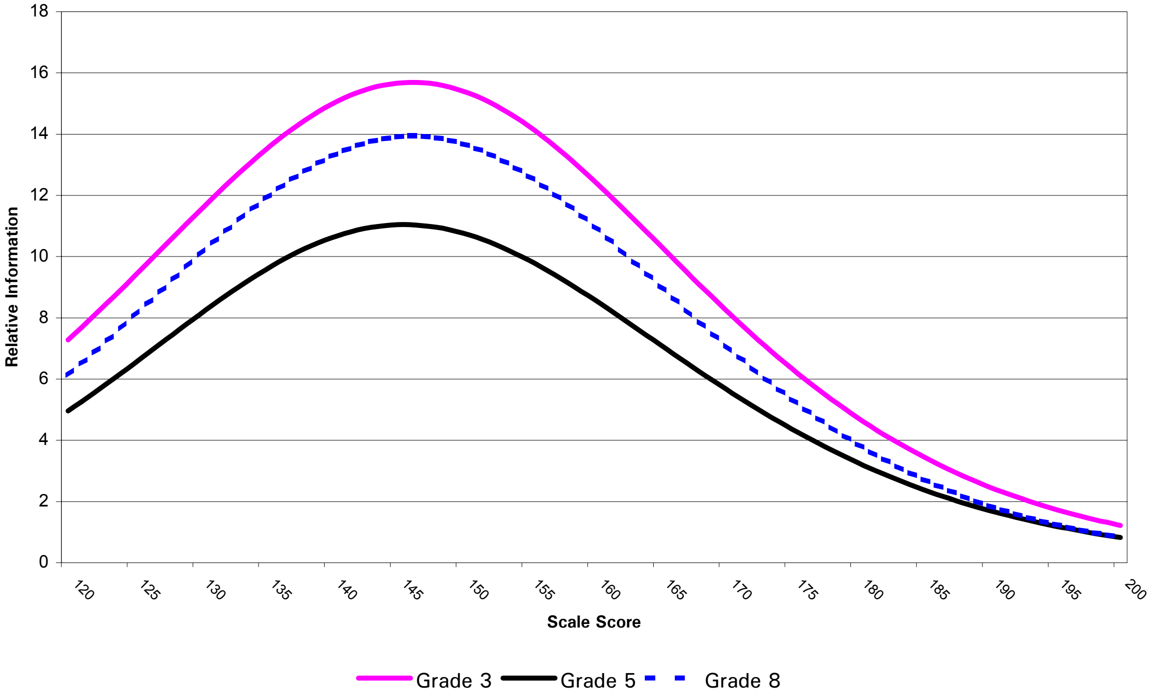


Figure 2.2
ISAT Mathematics Test Information Functions

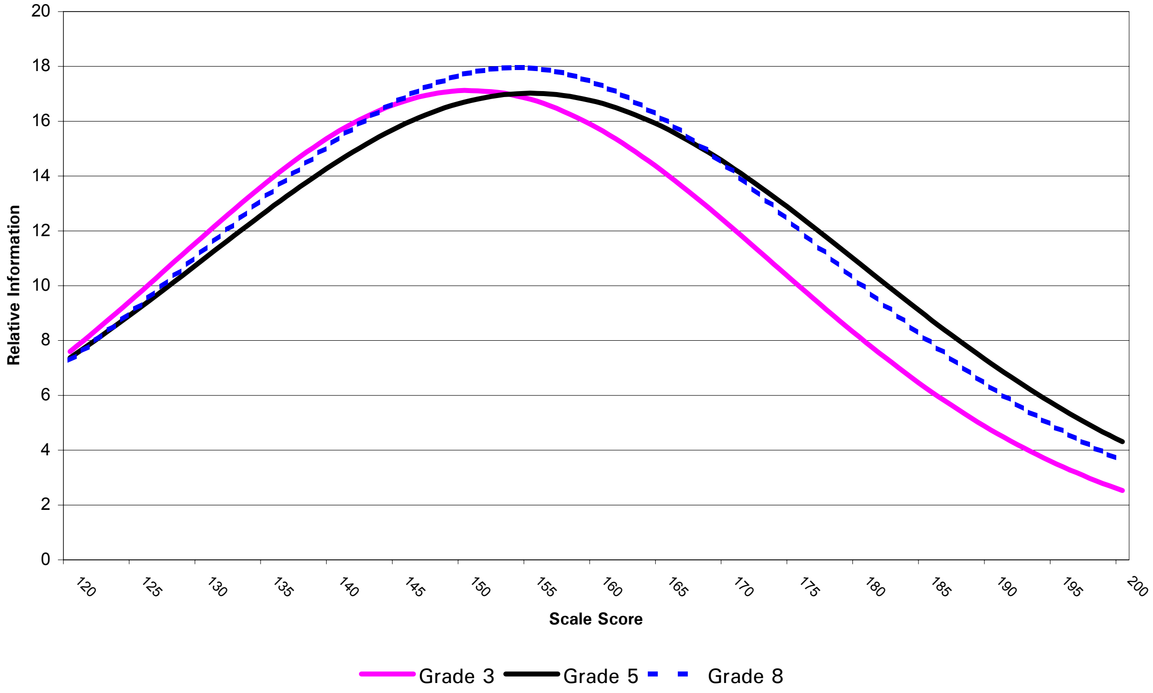
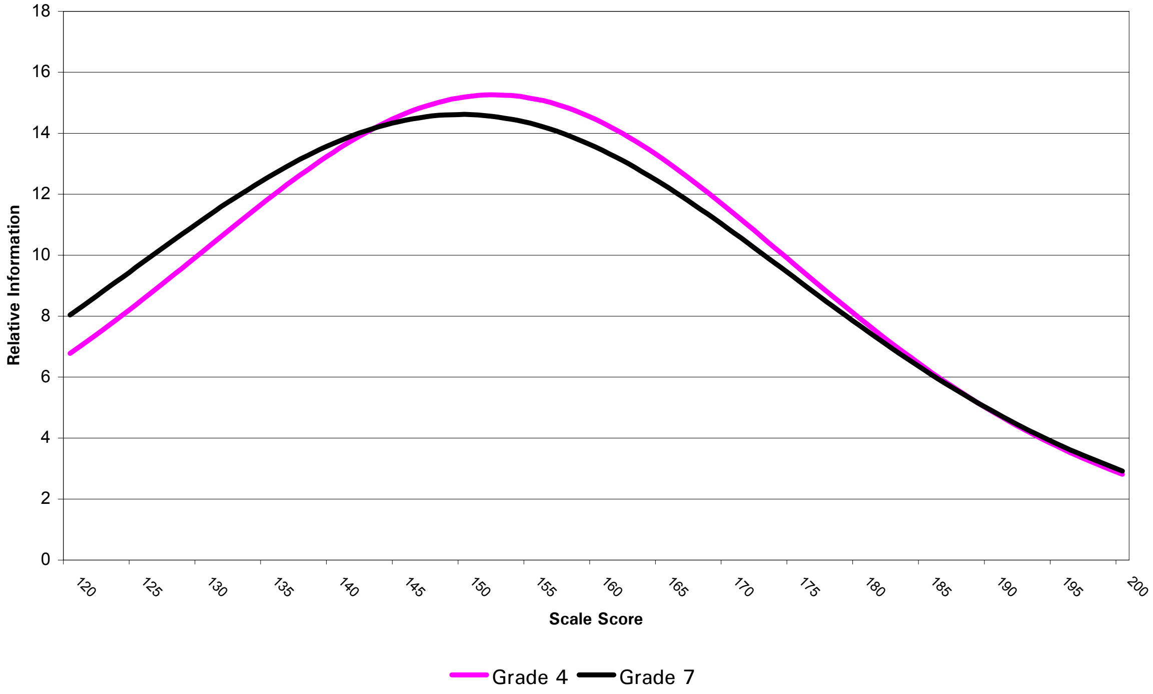


Figure 2.3
ISAT Science Test Information Functions



A second way of evaluating precision from the IRT perspective is in terms of how well the test as a whole separates people. 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). These values are reported in Table 2.2. Person separation values of 3 and above indicate a high degree of measurement precision. As the table indicates, the ISAT reading, mathematics, and science tests show consistently high levels of test precision across all the grade levels tested. Person separation values for the reading and mathematics tests are exceptionally high.

Table 2.2
Person Separation Values for the ISAT Tests

	Reading	Mathematics
Grade 3	3.38	3.44
Grade 5	2.94	3.58
Grade 8	3.05	3.93
	Science	
Grade 4	3.03	
Grade 7	2.97	

Reliability of the Extended-Response Scores

When scores integrate constructed response items they are affected by other sources of variance, particularly readers (raters), since different readers evaluate different students and items.

Interrater Agreement. Interrater agreement evaluates the consistency of scores assigned to the same response by different readers. For the constructed response items, interrater agreement was monitored daily, and two readers independently scored 10% of the items across grades.

For the reading test, raters provided a single score for each extended-response item, while extended-response items in the mathematics test were scored for knowledge, strategy, and explanation. Tables 2.3 and 2.4 present interrater agreement statistics for extended responses in reading and mathematics, respectively. The results for the interrater agreement on double-scored items exceeded the minimum acceptable level of agreement (90% agreement within one point). Scores across raters agree within one point at least 92% of the time.

Table 2.3
Interrater Agreement for Reading Extended-Response Items

	% Exact Agreement	% Adjacent Agreement	% Exact + Adjacent
Grade 3			
(N = 12,514)			
Item 1	60	37	97
Item 2	64	34	98
Grade 5			
(N = 12,949)			
Item 1	70	30	99
Grade 8			
(N = 14,123)			
Item 1	74	26	100
Item 2	74	26	100

Table 2.4
Interrater Agreement for Mathematics Extended-Response Items

	Score	% Exact Agreement	% Adjacent Agreement	% Exact + Adjacent
Grade 3				
Task 1	Knowledge	91	6	97
(N = 10,315)	Strategy	84	12	96
	Explanation	60	32	92
Task 2	Knowledge	80	16	96
	Strategy	70	22	92
	Explanation	55	38	93
Grade 5				
Task 1	Knowledge	85	12	97
(N = 14,059)	Strategy	82	13	95
	Explanation	59	29	88
Task 2	Knowledge	88	8	96
	Strategy	90	5	95
	Explanation	67	27	94
Grade 8				
Task 1	Knowledge	86	13	99
(N = 14,646)	Strategy	86	13	99
	Explanation	68	27	95
Task 2	Knowledge	89	9	98
	Strategy	83	13	96
	Explanation	62	33	96

Agreement with Validation Papers. In addition to agreement across raters, scores are checked against a standard, or "validation," set of responses. The Validation Committee assigns the scores for these papers. Item packets, each containing 10 essays, were circulated among the readers. Responses for these check sets were chosen to represent a range of score points in all categories.

Readers encountered the validation packets at random intervals throughout the scoring, and some encountered several packets during the scoring process. Readers were unaware of the scores assigned to the papers by the committee. The extent of agreement between a reader's scores and the scores assigned to the papers was calculated every day during the scoring and shared with the readers. This process allowed for the monitoring of reader scoring. For the reading test, raters provided a single score for the extended-response item, while extended-response items in the mathematics test were scored for knowledge, strategy, and explanation. Tables 2.5 and 2.6 present agreement with validation papers for extended responses in reading and mathematics, respectively.

Table 2.5
Agreement with Validation Papers for Reading Extended-Response Items

	% Exact Agreement	% Adjacent Agreement	% Exact + Adjacent
Grade 3			
Item 1	75	25	100
(N = 1,152)			
Item 2	72	26	98
(N = 1,152)			
Grade 5			
Item 1[1]*			
Item 2			
(N = 1,200)	86	14	100
Grade 8			
Item 1	78	21	99
(N = 1,320)			
Item 2	82	18	100
(N = 1,320)			

[1] Prior to the administration of the grade 5 reading test, passage 1 on the test was inadvertently used in training workshops with teachers from around the state. Because of this security breach and to avoid disadvantaging any student or school, answers to questions associated with passage 1 were not used to compute grade 5 reading scores. The questions from passage 1 included multiple-choice questions and one extended-response question. Grade 5 reading results were computed with the questions from the remaining three reading passages on the test, which included multiple-choice questions and one other extended-response question.

Table 2.6
Agreement with Validation Papers for Mathematics Extended-Response Items

	Score	% Exact Agreement	% Adjacent Agreement	% Exact + Adjacent
Grade 3				
Task 1	Knowledge	95	4	99
(N = 1,475)	Strategy	88	11	99
	Explanation	79	19	98
Task 2	Knowledge	95	5	100
(N = 1,475)	Strategy	88	8	96
	Explanation	65	30	95
Grade 5				
Task 1	Knowledge	94	6	100
(N = 1645)	Strategy	95	4	99
	Explanation	79	18	97
Task 2	Knowledge	97	2	99
(N = 1645)	Strategy	98	1	99
	Explanation	77	19	96
Grade 8				
Task 1	Knowledge	95	4	99
(N = 1,415)	Strategy	95	4	99
	Explanation	76	23	99
Task 2	Knowledge	84	16	100
(N = 1,415)	Strategy	92	7	99
	Explanation	71	28	99

Reliability of the Performance Category Decisions

Students' ISAT scores are reported relative to four performance categories: Academic Warning, Below Standards, Meets Standards, and Exceeds Standards. Sets of score cutoffs were developed for each learning area and each grade. The development of the score cutoffs that define these categories is fully documented in separate publications available from ISBE (*Performance Levels for the Illinois Standards Achievement Tests: Reading, Mathematics, Writing* and *Performance Levels for the Illinois Standards Achievement Tests: Science, Social Science*). However, the process may be briefly described as follows.

Prior to the meetings of the standard-setting panels themselves, which took place during April 1999 (reading, mathematics) and April 2000 (science), ISBE convened committees of curriculum experts to develop concrete descriptions of student knowledge and skill levels that define the specific performance categories. Educators throughout Illinois extensively reviewed these descriptions.

Panels of recognized subject matter experts convened in Springfield to translate the verbal descriptions into cut scores on the ISAT tests (i.e., scores that define the boundaries between categories). Panelists were drawn from a pool of educators who had specific knowledge of student performance at the grade levels being assessed by ISAT and experience in assessing students at those grade levels. Panelists were selected to be broadly representative of the geographic and ethnic diversity of Illinois' public school system. A total of 170 educators participated in the standard-setting process. The distribution of educators across learning areas was as follows: mathematics—56; reading—52; science—30.

A procedure originally proposed by Angoff is one of the most frequently used methods for determining cut scores when multiple-choice test scores are used. It can be most simply described as a focused, judgmental process by knowledgeable content experts. The basic Angoff procedure fit the format of the ISAT reading, mathematics, and science tests.

In the most frequent application of the Angoff method (e.g., to establish a pass-fail standard), panelists are asked to examine an item and decide what proportion of minimally competent individuals will answer the question correctly. With respect to the ISAT, however, instead of being asked about minimally competent students, panelists were asked to indicate what percentage of three groups of students—those who were just above the Academic Warning/Below Standards boundary, those who were just above the Below Standards/Meets Standards boundary, and those who were just above the Meets Standards/Exceeds Standards boundary—would answer the question correctly. The ratings were made sequentially rather than simultaneously (i.e., panelists made all judgments relative to one cut score before moving to the next cut score). Item performance statistics were provided to help panelists anchor their ratings. The cutoff scores that resulted are shown in Table 2.7. Results of applying these cutoffs to the 2005 test population are shown later in Section 4.

The reliabilities of such classifications, which are criterion-referenced, are related to the reliabilities of the tests on which they are based, but they are not equivalent to the test reliabilities, which are based on norm-referenced measurement. Glaser (1963) was among the first to draw attention to this distinction, and Feldt and Brennan (1989) extensively reviewed the topic.

Table 2.7
ISAT Cutoffs for Each Performance Level

READING	Academic Warning	Below Standards	Meets Standards	Exceeds Standards
03	120-137	138-155	156-173	174-200
05	120-129	130-155	156-170	171-200
08	120-128	129-151	152-172	173-200
MATHEMATICS	Academic Warning	Below Standards	Meets Standards	Exceeds Standards
03	120-141	142-152	153-172	173-200
05	120-137	138-157	158-190	191-200
08	120-137	138-161	162-184	185-200
SCIENCE	Academic Warning	Below Standards	Meets Standards	Exceeds Standards
04	120-138	139-153	154-178	179-200
07	120-141	142-150	151-174	175-200

As Feldt and Brennan (1989, p. 140) point out, approaches to the development of reliability coefficients for criterion-referenced interpretations of test scores have been based either on squared-error loss or threshold loss. It is threshold loss, which evaluates the consistency with which people are consistently classified with respect to a criterion, that is of greater concern here. Specifically, the issue is how consistently do tests classify students with respect to the performance standards?

Two threshold-loss coefficients have been developed: *p*, the proportion of persons consistently classified on two parallel tests, and *k* (kappa), which corrects *p* for the proportion of consistent classifications that would be expected by chance. Because scores on classically parallel tests are rarely available in practice, methods have been developed to estimate these values from a single test (Subkoviak, 1984). An approach proposed by Peng and Subkoviak (1980) was applied to the performance classifications made on the basis of the tests.

Table 2.8 presents the values for *p, k,* and $p_{miss,}$ the expected proportion of inconsistent decisions, which is simply (1 – *p*). In interpreting the first two indexes, Feldt and Brennan (1989) suggest that *p* reflects the *consistency of decisions* made about examinees, whereas *k*, since it is corrected for chance, reflects the *contribution of the test* to the consistency of the decision.

Overall, the values suggest that decisions made with respect to the student performance classifications would be very consistent. Note that the *p* and *k* values are calculated for the complete test population. Values for other test populations (e.g., IEP students alone, non-IEP students only) may differ.

Table 2.8
Reliability of Student Performance Decisions Based on Test Scores

		Academic Warning/Below Standards			Below Standards/Meets Standards			Meets Standards/Exceeds Standards		
Area	Grade	p	kappa	p_{miss}	p	kappa	p_{miss}	p	kappa	p_{miss}
Reading	3	0.976	0.742	0.024	0.910	0.795	0.090	0.924	0.783	0.076
	5	0.982	0.599	0.018	0.862	0.713	0.138	0.892	0.692	0.108
	8	0.986	0.357	0.014	0.880	0.698	0.120	0.926	0.662	0.074
Mathematics	3	0.976	0.742	0.024	0.936	0.774	0.064	0.908	0.796	0.092
	5	0.974	0.625	0.026	0.944	0.762	0.056	0.904	0.797	0.096
	8	0.972	0.731	0.028	0.932	0.774	0.068	0.916	0.789	0.084
Science	4	0.964	0.614	0.036	0.884	0.697	0.116	0.914	0.678	0.086
	7	0.942	0.642	0.058	0.862	0.624	0.138	0.902	0.691	0.098
AVERAGE		0.953	0.628	0.047	0.873	0.660	0.127	0.908	0.685	0.092

3. SCALING AND EQUATING PROCEDURES

ISAT reading, mathematics, and science scores are reported on a standard score scale. Individual student scores on this scale 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 of the first-year test population. Every unit on the scale represents 1/15 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 for each grade are 160 and 15. Results in subsequent years are equated to the base-year scale. The scaling constants used to transform the Rasch proficiency estimates to the reporting scale are shown in Table 3.1.

Table 3.1
ISAT Scaling Constants

	Slope	Intercept
Reading		
Grade 3	12.6428	146.2066
Grade 5	12.0100	144.7660
Grade 8	11.2280	141.7730
Mathematics		
Grade 3	13.5122	147.6910
Grade 5	14.9686	153.4644
Grade 8	14.7578	146.7806
Science		
Grade 4	15.3781	152.4255
Grade 7	15.9209	152.4527

Because test items change each year, raw scores (i.e., number or percent correct scores) will not always have the same meaning or represent the same level of proficiency. Without equating, each administration of a test with different items would lead to a new reporting scale, independent of that used previously. It would still be possible to measure relative performance, but it would not be possible to indicate growth across years for schools, districts, or the state. The equating process makes longitudinal comparisons possible.

The statistical fit of the one-parameter logistic (1PL) or Rasch model to the ISAT multiple-choice tests has been previously examined 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.

The equating procedures may be summarized as follows. Each test contains a sufficient number of items that have been previously administered to provide a reliable and content-representative equating link. During calibration of the new tests, item difficulties for these linking items are set to their historical values. By estimating values for the remaining items under this constraint, difficulty values for the remaining items are expressed on the existing scale. That is, the proficiency (theta) scale that results from the constrained calibration run is equated to the existing scale. The final step in the procedure is to apply equations that transform values on the proficiency scale to their corresponding ISAT scale score values. These equations were originally developed during the first year of equating and are then applied in each subsequent year of equating.

The logic of the equating procedure rests on certain assumptions. The most important is that the items used for linking stay the same in the two tests. During the assembly of tests, items that will be used for equating are placed exactly at or very near the location in the booklet where they previously appeared to minimize effects from positional differences. Differences between the anchored difficulties and the best-fit values are examined to ensure that no unusually large differences exist that would strain the equivalence assumption.

The equating analyses are conducted on samples of approximately 16,000 drawn from the total test population. A 1/nth selection results in a sample that has characteristics essentially identical with that of the total population.

Successive years' test forms, which have different items, are equated so that test scores will remain comparable across administrations. Each new test form contains a sufficient number of items that have been previously administered to provide a reliable and content-representative equating link. During calibration of the new tests, item difficulties for these linking items are set to their historical values. By estimating values for the remaining items under this constraint, difficulty values for the remaining items are automatically adjusted to the existing scale. The final step in the procedure is to apply equations that transform values on the proficiency scale to their corresponding scale score values. These equations were developed during the first year of testing.

Tables 3.2 through 3.4 show results of the Rasch calibration and equating procedures for reading. Column 1 of each table shows the item number within the test booklet. Column 2 shows the Rasch difficulties resulting from an anchored (constrained) calibration of the test. 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.

Tables 3.5 through 3.7 show similar information for the mathematics tests. Tables 3.8 and 3.9 present information for the science tests.

Table 3.2
Results of the Equating Process–Reading Grade 3

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
1	-1.37	0.03	0.96	0.82	0.35
2	0.25	0.02	1.06	1.06	0.41
3	-0.83	0.02	1.04	1.00	0.34
4	-0.47	0.02	1.03	1.08	0.38
5	-1.20	0.03	0.98	0.92	0.35
6	-0.56	0.02	1.01	0.98	0.39
7	-0.76	0.02	0.99	1.05	0.37
8	-1.13	0.03	0.87	0.69	0.44
9	-1.17	0.03	0.99	0.97	0.34
10	0.32	0.02	1.11	1.15	0.37
11	0.54	0.02	1.11	1.16	0.38
12	0.32	0.02	1.20	1.26	0.31
13	-0.26	0.02	1.17	1.14	0.31
14	0.08	0.02	1.12	1.14	0.36
15	-0.35	0.02	0.94	0.85	0.46
16	0.06	0.02	0.98	0.98	0.45
17	0.60	0.02	0.93	0.88	0.51
18	0.46	0.02	0.97	0.99	0.48
19	0.41	0.02	1.01	1.00	0.45
20	1.00	0.02	1.13	1.23	0.37
21	-0.94	0.02	0.87	0.66	0.46
22	0.23	0.02	1.09	1.15	0.38
23	-0.09	0.02	1.05	1.06	0.39
24	-0.21	0.02	0.86	0.73	0.53
25	-0.17	0.02	0.90	0.78	0.50
26	0.08	0.02	0.94	0.89	0.49
27	0.44	0.02	1.10	1.13	0.39
28	0.77	0.02	1.13	1.17	0.38
29	0.05	0.02	0.96	0.90	0.47
30	-0.14	0.02	0.97	0.97	0.44
31	-0.21	0.02	1.13	1.17	0.43
32	0.18	0.02	1.26	1.43	0.33
33	0.03	0.02	1.02	1.03	0.41
34	-0.36	0.02	1.00	0.98	0.43
35	0.27	0.02	0.92	0.82	0.51
36	-0.16	0.02	0.96	0.93	0.46
37	-0.10	0.02	1.04	1.10	0.44
38	-0.31	0.02	0.89	0.80	0.46
39	0.17	0.02	1.03	1.01	0.43
40	0.60	0.02	0.98	0.99	0.47
41	0.33	0.02	1.05	1.03	0.50
42	0.20	0.02	1.10	1.12	0.44
43	-0.21	0.02	0.98	0.99	0.46
44	1.07	0.02	1.07	1.10	0.42
45	-0.10	0.02	0.88	0.78	0.53
46	-0.02	0.02	1.12	1.16	0.42
47	-0.66	0.02	0.90	0.80	0.46
48	-2.03	0.03	0.87	0.69	0.32
49	-0.75	0.02	0.92	0.84	0.44

Table 3.2 (continued)

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
50	0.81	0.02	1.15	1.19	0.37
51	-0.94	0.02	0.84	0.79	0.43
52	-0.67	0.02	0.84	0.76	0.47
53	1.06	0.02	0.99	1.00	0.48
54	0.36	0.02	0.96	0.92	0.47
55	-0.39	0.02	0.76	0.59	0.56
56	0.11	0.02	0.90	0.82	0.52
57	-0.79	0.02	0.87	0.78	0.46
58	0.02	0.02	0.96	0.93	0.48
59	0.11	0.02	0.85	0.74	0.55
60	1.24	0.02	1.08	1.15	0.41
61	0.49	0.02	1.06	1.08	0.42
62	0.25	0.02	0.83	0.72	0.56
63	0.74	0.02	1.01	1.00	0.45
64	0.42	0.02	1.08	1.15	0.40
65	0.32	0.02	0.92	0.90	0.51
66	0.25	0.02	0.75	0.63	0.47
67	-0.76	0.02	0.94	0.97	0.36
68*	1.99	0.01	0.90	0.92	0.57
69*	2.54	0.01	1.12	1.13	0.53

* Extended-response item.

Table 3.3
Results of the Equating Process–Reading Grade 5

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
1	-1.26	0.03	1.03	1.29	0.28
2	0.31	0.02	0.96	0.90	0.48
3	-0.72	0.02	0.99	1.10	0.38
4	-0.15	0.02	0.85	0.74	0.54
5	-0.33	0.02	1.07	1.20	0.34
6	-0.54	0.02	0.89	0.75	0.49
7	-1.01	0.02	0.82	0.58	0.50
8	-0.36	0.02	0.85	0.75	0.53
9	0.24	0.02	0.91	0.88	0.51
10	1.72	0.02	1.03	1.14	0.41
11	0.70	0.02	0.98	1.01	0.47
12	0.14	0.02	0.88	0.82	0.53
13	0.70	0.02	0.92	0.90	0.51
14	-0.21	0.02	0.83	0.69	0.56
15	-0.57	0.02	0.96	1.03	0.41
16	0.22	0.02	1.08	1.09	0.38
17	1.65	0.02	1.05	1.17	0.39
18	0.80	0.02	1.32	1.50	0.19
19	-0.17	0.02	0.97	0.90	0.45
20	0.90	0.02	1.10	1.11	0.38
21	-0.42	0.02	0.93	0.82	0.46
22	0.14	0.02	0.92	0.86	0.50
23	-0.12	0.02	0.90	0.80	0.50
24	-0.68	0.02	0.91	0.84	0.45
25	0.26	0.02	1.05	1.06	0.41
26	-0.44	0.02	0.90	0.80	0.49
27	-0.36	0.02	0.93	0.86	0.47
28	-0.28	0.02	0.94	0.85	0.46
29	0.45	0.02	0.94	0.90	0.50
30	1.04	0.02	1.23	1.33	0.27
31	-1.23	0.03	0.85	0.62	0.46
32	-1.08	0.02	0.87	0.71	0.45
33	0.89	0.02	1.30	1.42	0.22
34	-0.69	0.02	0.83	0.69	0.51
35	0.35	0.02	0.98	0.92	0.50
36	0.55	0.02	1.07	1.10	0.37
37	1.42	0.02	1.11	1.20	0.35
38	0.25	0.02	1.22	1.36	0.28
39	-0.47	0.02	1.02	1.17	0.34
40	0.21	0.02	1.20	1.26	0.30
41	0.17	0.02	1.20	1.36	0.29
42	-0.63	0.02	0.91	0.82	0.45
43	2.11	0.02	1.12	1.34	0.33
44	-0.18	0.02	1.04	1.06	0.40
45	0.53	0.02	1.12	1.17	0.36
46	-0.24	0.02	0.98	0.92	0.42
47	-0.14	0.02	0.91	0.83	0.49
48	1.13	0.02	0.98	1.01	0.47
49	0.35	0.02	0.93	0.90	0.49

Table 3.3 (continued)

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
50*	2.96	0.01	0.91	0.91	0.59

* Extended-response item.

Table 3.4
Results of the Equating Process–Reading Grade 8

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
1	0.85	0.02	0.98	0.97	0.42
2	-0.14	0.02	0.95	0.88	0.34
3	0.33	0.02	0.89	0.78	0.55
4	0.43	0.02	0.93	0.88	0.47
5	1.56	0.02	1.19	1.31	0.25
6	1.71	0.02	1.17	1.28	0.31
7	0.87	0.02	1.00	0.99	0.44
8	0.17	0.02	0.91	0.84	0.45
9	-0.11	0.02	0.94	0.96	0.35
10	0.40	0.02	1.01	0.95	0.44
11	0.57	0.02	0.97	0.94	0.45
12	0.93	0.02	0.99	0.98	0.42
13	0.08	0.02	0.92	0.83	0.48
14	0.60	0.02	0.90	0.84	0.50
15	-0.08	0.02	0.90	0.78	0.51
16	1.01	0.02	0.97	0.98	0.46
17	-0.36	0.02	0.91	0.80	0.45
18	-0.16	0.02	0.92	0.82	0.43
19	-0.18	0.02	0.93	0.83	0.43
20	0.56	0.02	1.34	1.55	0.10
21	-1.03	0.03	0.92	0.82	0.37
22	-0.40	0.02	0.98	0.96	0.36
23	-0.56	0.02	0.98	0.98	0.35
24	-0.97	0.03	0.98	1.10	0.31
25	0.50	0.02	1.19	1.45	0.21
26	0.72	0.02	1.03	1.02	0.39
27	-0.50	0.02	1.00	1.18	0.32
28	1.42	0.02	1.02	1.06	0.41
29	-1.03	0.03	0.95	0.94	0.34
30	-0.36	0.02	0.99	1.08	0.34
31	-1.02	0.03	0.88	0.69	0.42
32	1.85	0.02	1.02	1.09	0.39
33	0.41	0.02	0.91	0.83	0.49
34	0.32	0.02	1.05	1.07	0.35
35	1.27	0.02	1.07	1.12	0.36
36	2.70	0.02	1.10	1.45	0.26
37	1.19	0.02	1.14	1.22	0.30
38	1.37	0.02	1.05	1.07	0.39
39	0.37	0.02	0.92	0.85	0.48
40	0.48	0.02	1.08	1.20	0.33
41	1.80	0.02	0.99	1.03	0.43
42	0.76	0.02	0.98	0.96	0.43
43	0.02	0.02	0.95	0.89	0.43
44	-0.27	0.02	0.87	0.72	0.48
45	0.84	0.02	0.98	0.98	0.43
46	0.79	0.02	0.91	0.85	0.50
47	1.23	0.02	1.02	1.05	0.40
48	0.16	0.02	1.04	1.04	0.40
49	-1.56	0.03	1.02	0.70	0.41

Table 3.4 (continued)

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
50	-0.59	0.02	0.98	0.84	0.45
51	0.90	0.02	1.12	1.13	0.33
52	-0.74	0.02	0.98	0.85	0.42
53	0.25	0.02	1.04	1.06	0.36
54	0.99	0.02	1.14	1.23	0.31
55	-1.14	0.03	0.97	0.68	0.47
56	-1.71	0.03	0.95	0.56	0.41
57	-0.43	0.02	0.99	1.04	0.36
58	1.50	0.02	1.09	1.16	0.34
59	0.47	0.02	1.04	1.07	0.39
60	0.95	0.02	1.01	1.02	0.43
61	-0.50	0.02	1.14	1.10	0.37
62	1.49	0.02	1.12	1.20	0.31
63	0.90	0.02	1.11	1.19	0.33
64	1.36	0.02	1.15	1.21	0.30
65	2.76	0.01	0.82	0.83	0.56
66	2.54	0.01	0.85	0.85	0.51

* Extended-response item.

Table 3.5
Results of the Equating Process–Mathematics Grade 3

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
1	-1.08	0.03	0.91	0.65	0.45
2	-1.28	0.03	0.88	0.63	0.40
3	0.23	0.02	0.90	0.81	0.50
4	-0.53	0.02	0.90	0.90	0.37
5	1.08	0.02	0.99	0.99	0.47
6	-0.72	0.02	0.94	0.84	0.40
7	-0.40	0.02	1.03	0.96	0.43
8	0.61	0.02	1.12	1.15	0.35
9	0.29	0.02	0.85	0.76	0.54
10	1.69	0.02	1.09	1.16	0.39
11	-0.19	0.02	0.88	0.88	0.48
12	-0.17	0.02	0.88	0.76	0.49
13	0.27	0.02	1.15	1.27	0.30
14	0.06	0.02	1.13	1.10	0.32
15	-0.17	0.02	1.13	1.26	0.35
16	-1.02	0.03	1.05	1.26	0.26
17	1.59	0.02	1.23	1.37	0.28
18	-0.80	0.02	0.97	0.87	0.43
19	-0.36	0.02	0.98	0.90	0.40
20	1.73	0.02	1.14	1.26	0.34
21	-0.99	0.03	1.12	1.58	0.19
22	0.19	0.02	0.98	0.98	0.43
23	0.84	0.02	0.93	0.88	0.49
24	1.48	0.02	0.90	0.87	0.53
25	1.44	0.02	0.94	0.94	0.50
26	-0.46	0.02	1.03	1.15	0.33
27	0.68	0.02	0.93	0.86	0.47
28	0.41	0.02	0.96	0.88	0.47
29	0.01	0.02	0.90	0.82	0.48
30	1.70	0.02	0.92	0.96	0.52
31	-0.10	0.02	0.89	0.81	0.48
32	1.35	0.02	1.06	1.09	0.41
33	-0.81	0.02	1.04	1.24	0.28
34	0.38	0.02	0.85	0.76	0.56
35	0.78	0.02	1.01	1.01	0.44
39	-0.46	0.02	0.77	0.69	0.31
40	-0.66	0.02	1.07	1.10	0.34
41	-0.76	0.02	0.95	0.92	0.38
42	0.49	0.02	1.05	1.08	0.46
43	-0.83	0.02	0.83	0.64	0.41
44	0.52	0.02	1.16	1.25	0.31
45	-0.24	0.02	0.95	0.98	0.29
46	0.54	0.02	1.01	1.03	0.44
47	1.05	0.02	1.03	1.04	0.44
48	1.43	0.02	1.00	1.00	0.46
49	-0.72	0.02	1.09	1.54	0.24

Table 3.5 (continued)

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
50	-0.29	0.02	0.82	0.72	0.47
51	0.82	0.02	0.92	0.89	0.51
52	-1.63	0.03	0.91	0.77	0.34
53	1.58	0.02	1.04	1.11	0.42
54	-1.21	0.03	1.01	0.81	0.40
55	1.43	0.02	0.96	0.96	0.49
56	-0.97	0.02	0.80	0.58	0.44
57	1.32	0.02	0.91	0.89	0.53
58	0.70	0.02	0.93	0.87	0.51
59	-0.12	0.02	0.89	0.77	0.50
60	0.81	0.02	1.07	1.08	0.37
61	1.15	0.02	1.07	1.10	0.40
62	-0.13	0.02	0.94	0.96	0.41
63	1.64	0.02	1.01	1.02	0.45
64	-0.15	0.02	1.06	0.98	0.36
65	0.56	0.02	1.10	1.11	0.38
66	-0.59	0.02	0.84	0.70	0.41
67	0.47	0.02	0.95	0.93	0.49
68	0.73	0.02	1.17	1.26	0.35
69	0.70	0.02	0.88	0.83	0.54
70	0.35	0.02	0.96	0.92	0.46
71	1.60	0.02	1.00	1.05	0.46
72	-0.89	0.02	0.83	0.56	0.47
73	0.57	0.02	0.92	0.86	0.50
ER1-K	-0.41	0.01	1.27	2.03	0.56
ER1-S	-0.45	0.01	1.13	1.28	0.55
ER1-E	0.88	0.01	1.27	1.32	0.52
ER2-K	0.22	0.01	1.56	1.88	0.56
ER2-S	0.18	0.01	1.25	1.46	0.63
ER2-E	1.21	0.01	1.20	1.22	0.57

Note: ER1, ER2 = Extended-response item; K = Knowledge score; S = Strategy score; E = Explanation score.

Table 3.6
Results of the Equating Process–Mathematics Grade 5

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
1	-2.53	0.04	0.95	0.81	0.28
2	-0.48	0.02	0.97	0.98	0.41
3	0.23	0.02	1.28	1.43	0.22
4	-0.87	0.02	0.98	0.96	0.39
5	-1.74	0.03	0.95	1.05	0.30
6	-1.05	0.02	0.90	0.84	0.41
7	0.40	0.02	1.02	1.02	0.43
8	0.07	0.02	1.02	1.04	0.43
9	0.66	0.02	1.09	1.11	0.39
10	0.67	0.02	1.07	1.08	0.41
11	0.80	0.02	0.95	0.95	0.50
12	1.51	0.02	0.96	1.02	0.49
13	-0.78	0.02	1.08	1.27	0.35
14	0.69	0.02	0.95	0.94	0.48
15	-0.40	0.02	0.95	1.01	0.46
16	-0.49	0.02	0.94	0.93	0.43
17	-1.32	0.02	0.97	1.18	0.27
18	1.03	0.02	0.97	0.97	0.49
19	0.81	0.02	0.91	0.88	0.53
20	-0.04	0.02	0.96	0.99	0.44
21	-0.32	0.02	0.97	1.00	0.42
22	-1.34	0.02	0.89	0.84	0.35
23	1.72	0.02	1.12	1.30	0.36
24	-0.04	0.02	1.02	1.13	0.39
25	0.89	0.02	0.97	0.96	0.49
26	-0.68	0.02	0.90	0.89	0.39
27	-0.39	0.02	0.93	0.88	0.45
28	-0.48	0.02	0.90	0.78	0.52
29	1.40	0.02	0.83	0.85	0.58
30	0.58	0.02	1.00	0.99	0.46
31	-0.90	0.02	1.01	1.01	0.36
32	0.86	0.02	1.02	1.04	0.44
33	0.97	0.02	1.01	1.03	0.46
34	1.16	0.02	0.86	0.87	0.57
35	0.36	0.02	1.03	1.05	0.42
39	-1.28	0.02	0.97	1.01	0.37
40	1.50	0.02	1.35	1.56	0.25
41	-1.84	0.03	0.91	0.89	0.32
42	1.30	0.02	1.02	1.07	0.44
43	-0.97	0.02	0.94	0.96	0.37
44	-1.54	0.03	0.96	1.13	0.27
45	0.24	0.02	0.86	0.78	0.54
46	1.45	0.02	1.00	1.05	0.45
47	-0.96	0.02	0.88	0.79	0.44
48	0.82	0.02	0.99	1.00	0.47
49	-0.20	0.02	1.10	1.09	0.35

Table 3.6 (continued)

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
50	0.79	0.02	1.02	1.02	0.45
51	0.58	0.02	1.08	1.09	0.41
52	0.16	0.02	1.02	1.03	0.47
53	-0.79	0.02	0.90	0.77	0.44
54	1.51	0.02	0.99	1.05	0.49
55	0.46	0.02	1.12	1.22	0.38
56	-0.14	0.02	0.92	0.83	0.53
57	0.58	0.02	0.86	0.82	0.55
58	0.80	0.02	1.05	1.06	0.43
59	-0.45	0.02	0.86	0.75	0.48
60	-0.18	0.02	1.00	1.02	0.41
61	-0.07	0.02	1.14	1.26	0.30
62	1.24	0.02	0.81	0.79	0.61
63	0.18	0.02	0.97	0.90	0.47
64	-0.18	0.02	1.10	1.14	0.40
65	-0.12	0.02	0.87	0.80	0.51
66	0.95	0.02	0.99	1.02	0.47
67	-0.84	0.02	0.82	0.79	0.38
68	-0.27	0.02	0.89	0.80	0.49
69	0.84	0.02	0.95	0.94	0.50
70	1.26	0.02	1.11	1.19	0.38
71	1.25	0.02	1.01	1.04	0.47
72	-0.21	0.02	0.92	0.97	0.46
73	0.36	0.02	0.97	0.99	0.46
ER1-K	-0.73	0.01	1.36	1.94	0.55
ER1-S	-0.63	0.01	1.32	1.93	0.56
ER1-E	-0.09	0.01	1.31	1.50	0.54
ER2-K	-0.24	0.01	1.44	1.97	0.59
ER2-S	-0.35	0.01	1.51	2.79	0.59
ER2-E	0.15	0.01	1.19	1.28	0.58

Note: ER1, ER2 = Extended-response item; K = Knowledge score; S = Strategy score; E = Explanation score.

Table 3.7
Results of the Equating Process–Mathematics Grade 8

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
1	-0.05	0.02	1.00	1.06	0.40
2	-1.09	0.02	1.09	1.70	0.25
3	1.17	0.02	0.99	0.97	0.51
4	-0.25	0.02	1.03	1.26	0.34
5	0.54	0.02	1.00	1.06	0.45
6	-0.22	0.02	0.95	1.07	0.42
7	1.93	0.02	0.96	1.00	0.54
8	-0.07	0.02	0.99	1.03	0.40
9	0.93	0.02	1.12	1.19	0.41
10	-0.46	0.02	0.89	0.81	0.44
11	-0.46	0.02	0.85	0.70	0.49
12	0.81	0.02	0.93	0.91	0.52
13	1.43	0.02	1.18	1.26	0.38
14	1.05	0.02	0.85	0.79	0.58
15	0.00	0.02	0.97	1.06	0.44
16	1.51	0.02	0.93	0.93	0.55
17	0.46	0.02	1.01	0.96	0.47
18	1.45	0.02	0.87	0.87	0.58
19	-0.97	0.02	0.88	0.76	0.42
20	0.77	0.02	0.92	0.87	0.53
21	0.26	0.02	1.00	1.02	0.44
22	1.85	0.02	1.02	1.07	0.49
23	0.76	0.02	0.90	0.85	0.54
24	0.81	0.02	0.93	0.87	0.52
25	0.67	0.02	0.93	0.92	0.51
26	0.97	0.02	0.92	0.92	0.55
27	0.53	0.02	0.82	0.76	0.59
28	0.77	0.02	1.12	1.20	0.39
29	0.35	0.02	0.97	0.91	0.46
30	0.01	0.02	0.93	0.86	0.48
31	-1.08	0.02	0.91	0.79	0.39
32	0.65	0.02	0.92	0.85	0.55
33	0.30	0.02	0.94	0.99	0.45
34	0.24	0.02	1.01	0.99	0.44
35	-1.14	0.02	0.89	0.79	0.39
39	-0.08	0.02	0.88	0.82	0.50
40	-1.07	0.02	0.96	1.02	0.35
41	0.40	0.02	1.05	1.22	0.42
42	0.24	0.02	0.88	0.78	0.53
43	-0.52	0.02	1.31	2.06	0.17
44	1.23	0.02	0.93	0.93	0.54
45	0.96	0.02	0.96	0.94	0.51
46	-0.69	0.02	1.04	1.38	0.27
47	0.61	0.02	0.79	0.73	0.62
48	0.55	0.02	1.05	1.01	0.43
49	1.39	0.02	1.14	1.19	0.41

Table 3.7 (continued)

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
50	0.69	0.02	0.85	0.78	0.57
51	0.92	0.02	1.09	1.11	0.43
52	1.32	0.02	1.13	1.17	0.42
53	0.57	0.02	1.06	1.11	0.43
54	1.04	0.02	0.96	0.94	0.52
55	0.91	0.02	1.17	1.27	0.37
56	1.04	0.02	0.99	0.95	0.50
57	1.04	0.02	1.25	1.33	0.33
58	1.48	0.02	1.11	1.15	0.44
59	0.57	0.02	1.05	1.12	0.43
60	0.98	0.02	1.04	1.09	0.46
61	0.76	0.02	0.97	0.96	0.49
62	0.53	0.02	0.84	0.75	0.57
63	1.19	0.02	1.17	1.30	0.38
64	1.09	0.02	0.97	0.98	0.51
65	0.78	0.02	0.88	0.84	0.56
66	1.36	0.02	1.03	1.05	0.48
67	0.54	0.02	0.94	0.92	0.50
68	-0.04	0.02	0.94	0.90	0.46
69	0.62	0.02	1.04	1.06	0.44
70	1.00	0.02	0.94	0.92	0.53
71	0.11	0.02	0.96	0.91	0.45
72	1.43	0.02	0.96	0.98	0.52
73	-0.01	0.02	0.86	0.79	0.42
ER1-K	0.42	0.01	0.96	0.95	0.71
ER1-S	0.46	0.01	1.13	1.17	0.68
ER1-E	-0.21	0.01	1.07	1.21	0.62
ER2-K	-0.15	0.01	1.34	1.71	0.61
ER2-S	-0.12	0.01	1.32	1.55	0.61
ER2-E	0.09	0.01	1.38	1.63	0.57

Note: ER1, ER2 = Extended-response item; K = Knowledge score; S = Strategy score; E = Explanation score.

Table 3.8
Results of the Scaling Process–Science Grade 4

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
1	-1.78	0.03	1.09	1.75	0.11
2	-0.57	0.02	1.05	1.13	0.31
3	0.01	0.02	1.00	1.00	0.40
4	-2.44	0.03	0.95	0.95	0.25
5	0.37	0.02	1.09	1.11	0.33
6	-0.21	0.02	1.05	1.09	0.34
7	-0.80	0.02	0.96	0.89	0.39
8	-0.28	0.02	0.92	0.85	0.45
9	0.20	0.02	0.98	0.98	0.42
10	0.05	0.02	1.02	1.00	0.32
11	-0.93	0.02	1.07	1.26	0.26
12	-0.40	0.02	0.95	0.88	0.43
13	0.45	0.02	1.11	1.14	0.32
14	-0.78	0.02	1.05	1.13	0.32
15	-0.29	0.02	0.97	0.95	0.36
16	-0.16	0.02	1.01	1.05	0.38
17	0.33	0.02	1.03	1.05	0.37
18	0.79	0.02	1.17	1.22	0.27
19	0.50	0.02	0.97	0.95	0.45
20	0.02	0.02	0.99	0.97	0.41
21	-0.06	0.02	0.96	0.92	0.44
22	0.84	0.02	0.96	0.96	0.45
23	-0.19	0.02	1.02	1.02	0.36
24	0.02	0.02	0.94	0.90	0.46
25	0.35	0.02	1.05	1.06	0.37
26	-1.61	0.03	0.89	0.74	0.40
27	0.59	0.02	1.03	1.04	0.39
28	-0.04	0.02	1.02	1.04	0.38
29	-0.45	0.02	0.89	0.83	0.47
30	0.86	0.02	1.01	1.02	0.41
31	-0.16	0.02	1.07	1.17	0.35
32	0.21	0.02	1.01	1.02	0.40
33	-0.65	0.02	0.91	0.84	0.44
34	0.21	0.02	0.99	0.99	0.44
35	-0.27	0.02	0.98	0.94	0.49
36	-0.13	0.02	1.03	1.07	0.36
37	0.72	0.02	1.15	1.22	0.28
38	-0.17	0.02	1.09	1.09	0.33
39	0.21	0.02	1.01	1.00	0.40
40	0.67	0.02	1.10	1.12	0.33
41	0.23	0.02	0.97	0.94	0.45
42	-0.02	0.02	0.93	0.89	0.44
43	-0.19	0.02	0.96	0.93	0.40
44	-0.41	0.02	0.91	0.87	0.38
45	0.81	0.02	1.14	1.19	0.29
46	-0.02	0.02	1.01	1.01	0.39
47	0.22	0.02	1.02	1.00	0.40
48	0.00	0.02	1.00	0.98	0.40
49	-0.14	0.02	0.85	0.77	0.53

Table 3.8 (continued)

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
50	-0.51	0.02	0.91	0.82	0.47
51	0.35	0.02	0.99	0.99	0.41
52	-0.65	0.02	0.99	0.99	0.40
53	0.49	0.02	0.99	0.98	0.42
54	-0.31	0.02	0.87	0.81	0.47
55	0.56	0.02	1.12	1.17	0.30
56	-0.19	0.02	0.96	0.93	0.42
57	-0.02	0.02	1.06	1.09	0.34
58	0.75	0.02	1.02	1.01	0.40
59	0.40	0.02	1.01	1.00	0.41
60	-0.23	0.02	0.88	0.86	0.43
61	-0.28	0.02	0.92	0.85	0.47
62	0.06	0.02	1.00	0.99	0.45
63	0.32	0.02	1.05	1.08	0.37
64	-0.47	0.02	0.84	0.73	0.52
65	-0.49	0.02	0.90	0.81	0.49

Table 3.9
Results of the Scaling Process–Science Grade 7

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
1	-1.79	0.03	0.94	0.84	0.34
2	-1.56	0.02	0.91	0.79	0.39
3	-0.95	0.02	0.92	0.83	0.39
4	0.66	0.02	0.96	0.94	0.46
5	0.22	0.02	0.96	0.93	0.45
6	-0.73	0.02	0.96	0.90	0.39
7	0.27	0.02	0.98	0.97	0.43
8	-0.99	0.02	1.00	0.91	0.42
9	-0.07	0.02	1.08	1.10	0.32
10	0.16	0.02	1.00	1.01	0.40
11	-0.11	0.02	0.98	0.97	0.40
12	-1.30	0.02	0.77	0.72	0.35
13	-0.93	0.02	0.99	0.92	0.37
14	-0.32	0.02	0.96	0.92	0.43
15	-0.58	0.02	0.95	0.99	0.38
16	-0.61	0.02	1.03	1.03	0.34
17	-1.17	0.02	0.96	1.06	0.27
18	0.60	0.02	1.03	1.05	0.38
19	0.08	0.02	1.07	1.08	0.34
20	-0.64	0.02	0.92	0.84	0.45
21	-0.53	0.02	0.94	0.89	0.43
22	-0.35	0.02	0.92	0.86	0.47
23	0.08	0.02	1.02	1.02	0.37
24	0.11	0.02	1.12	1.17	0.31
25	-0.22	0.02	0.95	0.91	0.44
26	0.34	0.02	0.96	0.95	0.44
27	0.17	0.02	1.15	1.17	0.27
28	0.32	0.02	1.09	1.12	0.33
29	-1.25	0.02	0.92	0.91	0.39
30	-1.75	0.03	0.89	0.69	0.40
31	-0.97	0.02	0.88	0.84	0.37
32	-0.89	0.02	0.81	0.70	0.46
33	-0.09	0.02	0.91	0.87	0.48
34	0.76	0.02	0.98	0.99	0.43
35	0.44	0.02	1.07	1.09	0.34
36	0.60	0.02	1.00	1.00	0.41
37	0.62	0.02	1.01	1.02	0.41
38	0.09	0.02	1.15	1.20	0.29
39	-1.01	0.02	0.98	0.95	0.36
40	-1.22	0.02	0.99	0.91	0.35
41	-0.14	0.02	1.03	1.10	0.34
42	1.04	0.02	1.04	1.08	0.37
43	1.02	0.02	1.06	1.10	0.37
44	1.12	0.02	1.11	1.19	0.30
45	0.58	0.02	1.05	1.08	0.37
46	-0.49	0.02	1.01	1.04	0.37
47	0.07	0.02	1.14	1.31	0.27
48	-0.38	0.02	0.92	0.88	0.43
49	-0.75	0.02	1.01	1.02	0.32

Table 3.9 (continued)

Item	Difficulty	S_{ed}	Infit	Outfit	r_{pb}
50	0.67	0.02	1.09	1.12	0.32
51	0.72	0.02	1.05	1.07	0.36
52	-0.93	0.02	0.87	0.77	0.47
53	0.05	0.02	0.94	0.93	0.46
54	0.28	0.02	0.98	0.98	0.43
55	0.41	0.02	0.99	0.99	0.42
56	-1.07	0.02	0.79	0.62	0.48
57	0.67	0.02	1.08	1.11	0.34
58	0.23	0.02	0.98	0.96	0.44
59	0.63	0.02	1.15	1.19	0.27
60	-0.29	0.02	1.03	1.05	0.39
61	-0.88	0.02	0.95	0.89	0.41
62	-0.72	0.02	0.85	0.72	0.52
63	-0.92	0.02	0.87	0.78	0.47
64	-0.04	0.02	1.04	1.02	0.39
65	0.57	0.02	1.04	1.07	0.37

4. RESULTS

Performance Relative to the Illinois Learning Standards

Table 4.1 shows the percentages of students by performance level and by grade for reading. The percentage of students falling into the Exceeds category is highest at 3rd grade. The percentage of students not meeting standards is highest at 5th grade. Overall, the percentage of students meeting (or exceeding) standards is highest at 8th grade.

Table 4.1
Percentages of Students by Grade Falling into Each Performance Level for ISAT Reading: 1999-2005

Grade/ Year	Academic Warning	Below Standards	Meets Standards	Exceeds Standards
3				
1999	8	31	44	17
2000	6	32	41	21
2001	7	31	43	19
2002	7	31	44	19
2003	8.2	29.9	40.1	21.9
2004	7.1	27.9	42.3	22.7
2005	6.6	26.7	45.1	21.5
5				
1999	1	38	37	24
2000	0	41	39	20
2001	1	40	34	25
2002	1	39	37	22
2003	1.0	38.6	37.3	23.1
2004	1.7	37.4	35.9	25.0
2005	1.8	38.3	40.4	19.4
8				
1999	1	27	54	18
2000	0	28	56	16
2001	1	34	56	10
2002	1	31	58	10
2003	0.5	35.8	54.0	9.7
2004	1.6	31.3	57.4	9.7
2005	0.7	26.6	61.3	11.5

Note: Because of rounding, the percentages in each row may not total exactly to 100%.

Table 4.2 provides additional information with respect to the reading test. It presents the average percent of items students answered correctly with respect to the standard sets that were previously described.

Table 4.2
Reading Average Percent Correct by Standard Sets

	Set					
Grade	1	2	3	4	5	6
03	70	69	73	69	68	78
05	72	67	69	70	63	–
08	72	65	74	67	65	–

Table 4.3 shows the percentages of students by performance level and by grade for mathematics. The percentage of students meeting state standards is highest for grade 3 students and lowest for grade 8 students. The percentage of students falling into the Exceeds category is much higher at grade 3 than at the other two grades.

Table 4.3
Percentages of Students by Grade Falling into Each Performance Level for ISAT Mathematics: 1999-2005

Grade/ Year	Academic Warning	Below Standards	Meets Standards	Exceeds Standards
3				
1999	12	20	47	21
2000	10	21	46	23
2001	8	18	46	28
2002	7	19	44	30
2003	6.8	17.4	44.6	31.1
2004	6.8	14.0	46.1	33.0
2005	5.3	15.4	45.2	34.1
5				
1999	6	39	53	3
2000	6	37	52	5
2001	4	34	55	6
2002	5	32	55	8
2003	3.5	28.1	58.6	9.7
2004	3.0	25.3	59.8	12.0
2005	3.2	23.6	60.8	12.4
8				
1999	5	52	36	7
2000	8	46	35	12
2001	7	42	37	13
2002	7	40	37	15
2003	6.3	40.6	37.6	15.5
2004	5.6	40.0	37.5	16.9
2005	5.9	39.7	37.4	16.9

Note: Because of rounding, the percentages in each row may not total exactly to 100%.

Table 4.4 presents the average percent of items students answered correctly with respect to the mathematics standard sets that were previously described.

Table 4.4
Mathematics Average Percent Correct by Standard Sets

	Set							
Grade	1	2	3	4	5	6	7	8
03	65	70	71	67	71	63	74	62
05	63	65	65	61	66	58	62	69
08	61	59	62	58	62	58	59	61

Table 4.5 shows the percentages of students by performance level and by grade for science.

Table 4.5
Percentages of Students by Grade Falling into Each Performance Level for ISAT Science: 2000-2005

Grade/ Year	Academic Warning	Below Standards	Meets Standards	Exceeds Standards
4				
2000	1	35	51	13
2001	8	26	54	11
2002	8	25	53	14
2003	7.0	26.5	52.2	14.3
2004	6.0	26.2	54.6	13.2
2005	5.0	23.6	55.1	16.3
7				
2000	12	16	54	18
2001	11	17	52	20
2002	10	17	56	17
2003	9.7	16.6	56.2	17.5
2004	10.4	15.2	57.8	16.6
2005	10.4	15.0	54.3	20.3

Note: Because of rounding, the percentages in each row may not total exactly to 100%.

Table 4.6 presents the average percent of items students answered correctly with respect to the science standards sets that were previously described.

Table 4.6
Science Average Percent Correct by Standard Sets

	Set				
Grade	1	2	3	4	5
04	70	64	63	59	65
07	69	69	62	52	69

Performance Relative to National Quarters

The legislation that authorized the development of ISAT required that reports provide national comparative data as a secondary reference point for evaluating school improvement efforts. Since the costs of obtaining nationally representative samples of students for each test would be prohibitively expensive, that mandate has been met by administering a nationally standardized achievement test along with ISAT to a sample of Illinois students. The two score distributions are then compared to identify points on the ISAT scale that correspond to the 25th, 50th, and 75th percentile performance levels for the national sample.

ISAT uses the Ninth Edition of the Stanford Achievement Tests (SAT9) for purposes of determining Illinois students' relative standing within the national population. Equipercentile methodology was used to equate scores on the two tests. In equipercentile equating, scores on two tests are assumed to be equivalent if they have the same percentile rank. For example, the SAT9 score that cuts off 10% of the equating sample is assumed to represent a level of proficiency equal to the ISAT score that cuts off 10% of the equating sample, even though the scores themselves may be quite different numerically.

Table 4.7 presents the ISAT scale score cutoffs that define the *upper limits* of national quartile categories 1, 2, and 3. These are shown as score ranges for each national quarter. For example, scale scores of 120 to 145 on the 4th-grade science test define Q1, the quartile that represents the lowest 25% of student performance nationally. Note that although the scale score cutoffs remain the same from year to year, the percentage of students in each category need not remain constant.

The results of applying these cutoffs to the 2005 assessment data are shown in Table 4.8.

Table 4.7
ISAT National Quarter Scale Score Cutoffs

READING	Q1	Q2	Q3	Q4
03	120-147	148-157	158-167	168-200
05	120-147	148-157	158-168	169-200
08	120-144	145-154	155-165	166-200
MATHEMATICS	Q1	Q2	Q3	Q4
03	120-145	146-155	156-166	167-200
05	120-146	147-156	157-166	167-200
08	120-144	145-154	155-164	165-200
SCIENCE	Q1	Q2	Q3	Q4
04	120-145	146-157	158-168	169-200
07	120-142	143-154	155-163	164-200

Table 4.8
Percentages of Students by Grade and Learning Area Falling into Each National Quartile: 1999-2005

READING	Q1	Q2	Q3	Q4
Grade/Year				
3				
1999	22	22	25	32
2000	21	21	25	33
2001	21	22	25	32
2002	21	21	26	33
2003	22	20	25	33
2004	19	20	26	35
2005	18	21	23	37
5				
1999	21	23	27	28
2000	21	26	28	25
2001	25	21	24	30
2002	23	23	26	28
2003	23	22	27	28
2004	22	23	27	28
2005	21	22	33	24

Table 4.8 (continued)

8				
1999	15	22	30	33
2000	13	24	33	30
2001	17	26	33	24
2002	17	23	34	25
2003	19	27	31	24
2004	16	24	35	25
2005	12	25	35	28

MATHEMATICS	Q1	Q2	Q3	Q4
Grade/Year				
3				
1999	19	21	28	32
2000	18	21	26	36
2001	14	19	25	42
2002	13	19	25	43
2003	12	18	25	44
2004	10	17	28	46
2005	9	18	27	47
5				
1999	20	22	24	33
2000	19	22	21	38
2001	17	19	21	42
2002	16	19	22	43
2003	13	17	21	49
2004	10	16	24	49
2005	11	15	22	53
8				
1999	15	25	25	35
2000	18	20	21	41
2001	17	19	18	45
2002	16	19	20	46
2003	16	17	18	48
2004	14	18	18	50
2005	15	18	19	48

Table 4.8 (continued)

SCIENCE	Q1	Q2	Q3	Q4
Grade/Year				
4				
2000	18	26	25	31
2001	19	23	27	30
2002	18	24	27	30
2003	18	25	25	32
2004	16	26	26	32
2005	13	25	25	37
7				
2000	14	24	22	41
2001	12	25	20	43
2002	12	25	23	41
2003	11	23	24	42
2004	12	23	23	42
2005	12	23	20	45

Note: Because of rounding, the percentages in each row may not total exactly to 100%.

Correlations Among Scale Scores

Correlations among the scale scores at each grade tested are presented in Table 4.9. Appendix A provides correlations among the standard sets. The sample sizes on which the correlations in Table 4.9 are based are also shown in Appendix A.

Table 4.9
Correlations Among ISAT Scale Scores

Grade 3	Reading	Mathematics
Reading	1.000	.782
Mathematics	.782	1.000

Grade 5	Reading	Mathematics
Reading	1.000	.753
Mathematics	.753	1.000

Grade 8	Reading	Mathematics
Reading	1.000	.749
Mathematics	.749	1.000

References

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement (3rd Edition)* (pp. 105-146). New York: Macmillan.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18,* 519-521.

Peng, C-Y, J., & Subkoviak, M. J. (1980). A note on Huynh's normal approximation procedure for estimating criterion-referenced reliability. *Journal of Educational Measurement, 17,* 359-368.

Subkoviak, M. J. (1984). Estimating the reliability of mastery/non-mastery classifications. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 267-291). Baltimore: Johns Hopkins Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago: Mesa.

APPENDIX A. SUPPLEMENTARY TABLES

Tables A.1 through A.3 present correlations among the various standard sets, goal, or feature scores presented in student, school, and district reports. The sample sizes for the various analyses are summarized below.

Reading: Grade 3	137,309
Reading: Grade 5	148,635
Reading: Grade 8	154,944
Mathematics: Grade 3	137,562
Mathematics: Grade 5	148,816
Mathematics: Grade 8	155,190
Science: Grade 4	144,479
Science: Grade 7	155,270

Table A.1
Correlations Among Reading Standard Sets

Grade 3	S1	S2	S3	S4	S5	S6
S1	1.000	.809	.845	.974	.789	.625
S2	.809	1.000	.831	.878	.737	.597
S3	.845	.831	1.000	.806	.657	.567
S4	.974	.878	.806	1.000	.745	.628
S5	.789	.737	.657	.745	1.000	.600
S6	.625	.597	.567	.628	.600	1.000

Grade 5	S1	S2	S3	S4	S5
S1	1.000	.731	.707	.890	.684
S2	.731	1.000	.895	.912	.756
S3	.707	.895	1.000	.769	.610
S4	.890	.912	.769	1.000	.676
S5	.684	.756	.610	.676	1.000

Grade 8	S1	S2	S3	S4	S5
S1	1.000	.772	.718	.940	.528
S2	.772	1.000	.875	.915	.610
S3	.718	.875	1.000	.763	.461
S4	.940	.915	.763	1.000	.538
S5	.528	.610	.461	.538	1.000

Table A.2
Correlations Among Mathematics Standard Sets

Grade 3	S1	S2	S3	S4	S5	S6	S7	S8
S1	1.000	.899	.817	.790	.761	.886	.819	.739
S2	.899	1.000	.735	.699	.700	.735	.746	.638
S3	.817	.735	1.000	.683	.659	.687	.679	.631
S4	.790	.699	.683	1.000	.814	.724	.670	.648
S5	.761	.700	.659	.814	1.000	.663	.652	.633
S6	.886	.735	.687	.724	.663	1.000	.675	.649
S7	.819	.746	.679	.670	.652	.675	1.000	.679
S8	.739	.638	.631	.648	.633	.649	.679	1.000

Grade 5	S1	S2	S3	S4	S5	S6	S7	S8
S1	1.000	.882	.803	.809	.780	.916	.907	.844
S2	.882	1.000	.761	.693	.717	.774	.781	.705
S3	.803	.761	1.000	.684	.664	.713	.742	.690
S4	.809	.693	.684	1.000	.852	.745	.706	.667
S5	.780	.717	.664	.852	1.000	.692	.704	.657
S6	.916	.774	.713	.745	.692	1.000	.800	.721
S7	.907	.781	.742	.706	.704	.800	1.000	.739
S8	.844	.705	.690	.667	.657	.721	.739	1.000

Grade 8	S1	S2	S3	S4	S5	S6	S7	S8
S1	1.000	.920	.844	.872	.860	.932	.882	.868
S2	.920	1.000	.781	.812	.825	.843	.785	.749
S3	.844	.781	1.000	.752	.761	.777	.709	.715
S4	.872	.812	.752	1.000	.795	.880	.744	.720
S5	.860	.825	.761	.795	1.000	.809	.732	.725
S6	.932	.843	.777	.880	.809	1.000	.805	.762
S7	.882	.785	.709	.744	.732	.805	1.000	.732
S8	.868	.749	.715	.720	.725	.762	.732	1.000

Table A.3
Correlations Among Science Standard Sets

Grade 4	S1	S2	S3	S4	S5
S1	1.000	.667	.627	.680	.687
S2	.667	1.000	.620	.689	.666
S3	.627	.620	1.000	.648	.622
S4	.680	.689	.648	1.000	.673
S5	.687	.666	.622	.673	1.000

Grade 7	S1	S2	S3	S4	S5
S1	1.000	.699	.641	.638	.685
S2	.699	1.000	.640	.647	.689
S3	.641	.640	1.000	.592	.620
S4	.638	.647	.592	1.000	.633
S5	.685	.689	.620	.633	1.000