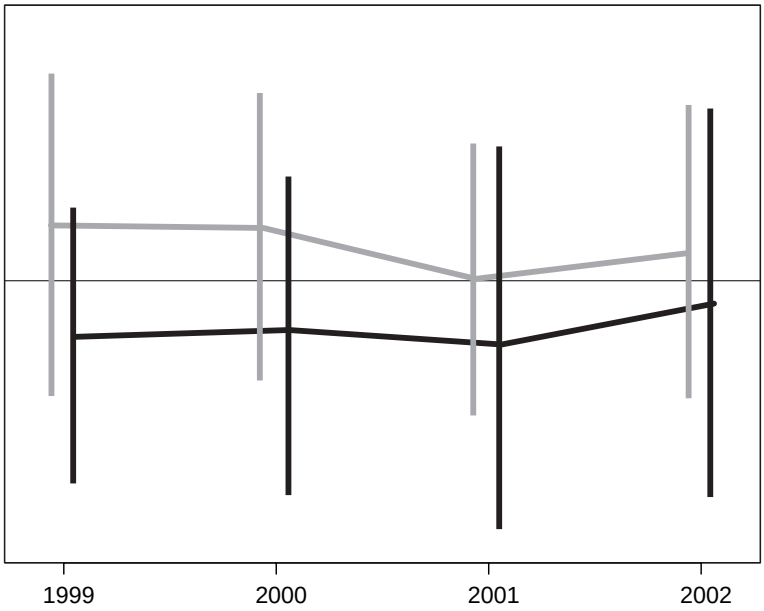


Research Data Brief

How Do They Compare?

ITBS and ISAT Reading and Mathematics in the Chicago Public Schools, 1999 to 2002

February 2003

John Q. Easton
Macarena Correa
Stuart Luppescu
Hye-Sook Park
Stephen Ponisciak
Todd Rosenkranz
Susan Sporte

Consortium on Chicago School Research

This is the fifth in a series of research data briefs designed to provide new data on a particular issue. As the name suggests, this is a short report focusing on a single topic. Because data briefs are not comprehensive studies, we limit our discussion on findings to summarizing the key results.

This data brief reflects the interpretations of the authors. Although the Consortium's Steering Committee provided technical advice and reviewed an earlier version, no formal endorsements by these individuals, their organizations, or the full Consortium should be assumed.

Overview

For several decades, the Iowa Tests of Basic Skills (ITBS) held a preeminent role in measuring student and school performance in the Chicago Public Schools (CPS). Local newspapers—both the dailies and neighborhood weeklies—traditionally printed ITBS results to summarize achievement levels in CPS schools. With the ending of social promotion and the introduction of school probation in the mid-to-late 1990's, the ITBS took on greater significance for individual students, schools, and teachers and principals. Schools were held accountable for their ITBS scores; they were placed on or off probation based on the percentage of students who scored at or above national norms. In addition, the test's use to determine which students needed to attend mandatory summer school and which were eligible for promotion to the next grade made it a high-stakes test.

No Child Left Behind (NCLB)—recent federal legislation that increases accountability for student academic performance—is changing the testing and assessment landscape across the country.[1] Among its provisions, NCLB requires all states to conduct annual achievement testing (based on state learning standards in mathematics and reading) of all students in grades three through eight by the 2005-06 school year. In planning to meet this schedule, the Illinois State Board of Education (ISBE) released a "State Assessment Proposal" with a timeline for projected activities.[2] By July 2003, ISBE hopes to award a contract for developing state assessments to comply with NCLB.

In the fall of 2002, CPS announced a new accountability system for judging schools' academic progress. In a deliberate effort to broaden the criteria for measuring student performance in schools, and to prepare for NCLB, the new accountability system included results from the Illinois Standards Achievement Test (ISAT), the state's elementary assessment since 1999. In the new system, two of the four statistical indicators upon which elementary schools are held accountable are based on the ISAT.[3] In the past, all CPS accountability was ITBS based.

These changes—ones that have already taken place and others that will happen soon—are occurring despite a relative lack of public knowledge in Chicago about the ISAT. With NCLB requiring such major changes in assessment, it is important to develop greater understanding of achievement testing in our schools. Because of such strong emphasis on the ITBS within CPS, the ISAT has received little of the public scrutiny that has been focused on the ITBS. In this data brief, we compare and contrast the two tests to bring some needed attention to the ISAT and, by doing so, perhaps assist the discussions about the next generation of assessments that must be put into place by the 2005-06 school year.

For More Detailed Information About:

ITBS

- Allotted time and number of questions for K-8 grade level tests
- Brief descriptions of content by subject
- Correlations and explanations of different test forms
- Scoring
- ITBS score conversions
- Updates for norms and scoring on the new ITBS test
- How CPS measures student gains
- Latest version of the ITBS
- Understanding the scores and scoring as used in CPS

ISAT

- Reading and mathematics tests and content
- Performance definitions and cut-points
- Sample test books for reading and mathematics with extended response questions
- Technical manuals
- Test guides
- Reliability estimates
- Technical analysis including validity, reliability, and test equating, done in comparison with the SAT-9 test
- Comparisons between the ITBS and the IGAP
- Purpose of the ISAT
- Test descriptions

Both

- Content covered in both assessments
- Sample problems

GO TO

www.consortium-chicago.org/publications/briefs/onlineresources.html

for online resources.

Differences between the ITBS and ISAT

What Are the Purposes of the Two Tests?

The ISAT and ITBS were created to serve two different but not mutually exclusive purposes. The ISAT was developed to measure the extent to which students in Illinois public schools have met the Illinois Learning Standards, which are specific, state-defined learning goals. The ISAT and other tests that are constructed to measure students' performance on specific subject matter content are called Criterion-Referenced Tests (CRTs). Results on CRTs indicate whether or not students demonstrate pre-determined levels of mastery regarding the specified subject matter. Its content is based on learning standards developed by Illinois educators. State educators and curriculum experts submit items for the ISAT. In addition, material considered for inclusion in future examinations is pilot tested each year.

The ITBS, on the other hand, is designed to compare students to a nationally representative group on widely acknowledged basic skills. Because the ITBS compares the performance of students to a national average, it is called a Norm-Referenced Test (NRT). Since the ITBS is administered in many school districts around the country, it is impossible to take into consideration curriculum standards that are specific to one particular district or state. Instead, the test is based on current instructional materials (i.e., textbooks; subject-matter content standards issued by professional organizations, such as the National Council of Teachers of Mathematics), guidance from curriculum specialists, and analysis of previous test forms to reflect widely used material and curricula. As its name indicates, the ITBS is intended to measure students' mastery of *basic* skills, with less emphasis on specialized abilities and higher-order skills. NRTs are scored to indicate student performance relative to other students rather than to the content material itself.

The ITBS is also designed to measure annual student learning growth. Because it has a continuous scale across all grades and is administered every year to nearly all CPS students in grades three through eight, it is used to measure student achievement growth. Use of the ISAT for measuring student growth is constrained by two factors: it is only administered to key "benchmark" grades (three, five, and eight) and there is not a unitary underlying scale across the grades.

Test Format Differences

In comparing these two tests, we start with their simple differences, such as how they are administered, how many questions they contain, and how the questions differ from each other. We later proceed to more complex issues, such as scoring and scaling.

The ITBS reading comprehension test is administered in two sessions for a total of 55 minutes (the first is 25 minutes and the second is 30 minutes). The ISAT reading test is administered in three sessions of 40 minutes each, for a total of 120 minutes (see Table 1). The ITBS mathematics test consists of three sub-tests. Students are permitted 30 minutes for each of the first two subtests and 20 minutes for the third, making a total of 80 minutes. The ISAT mathematics test includes three

Table 1. Number of Testing Sessions and Total Time

	Reading		Math	
	ITBS	ISAT	ITBS	ISAT
Number of sessions	2	3	3	3
Minutes per session	25, 30	40	30, 30, 20	40
Total time	**55**	**120**	**80**	**120**

sections, each of which lasts 40 minutes, for a total of 120 minutes. In short, the ISAT takes longer to complete than the ITBS with more than twice as much time in reading, and half again as much in mathematics.

One reason why the time requirements for the two tests differ is that the number and type of their items differ. Whereas the ITBS is made up entirely of multiple-choice questions, the ISAT includes extended-response questions as well. These questions require students to write in responses to prompts about the reading passages and to explain how they completed certain mathematics problems. The ISAT has two extended-response questions on both the reading and mathematics sections, except for third-grade reading, where there is only one.

The general pattern is that on the ISAT there are more reading items and fewer mathematics items than on the ITBS, as shown in Table2 below.

Table 2. Number of Test Questions

	Reading		Math	
	ITBS	ISAT	ITBS	ISAT
Grade 3	37	68	90	72
Grade 5	43	67	113	72
Grade 8	52	67	135	72

The ITBS reading comprehension test contains eight passages followed by between 6 and 10 multiple-choice questions based on the passage. There are three reading passages at all three ISAT grades, each of which is followed by between 15 and 20 multiple-choice questions per passage. For grades five and eight, one of the three passages is typically longer than the other two.

Content Differences: Reading

The goals and foci of the ITBS and ISAT are slightly different.[4] The ITBS reading comprehension test measures skills important in reading comprehension. According to the *Iowa Tests Interpretive Guide for Teachers and Counselors*, "The primary focus of the test and of the use of scores is on skills."[5] Because of its relationship to the Illinois Learning Standards, the goal of the ISAT is somewhat broader. According to the *Illinois Standards Achievement Test Sample Reading Materials 2002,* the ISAT reading comprehension passages reflect the most frequent purposes of reading: "reading to gain information and reading for literary experience."[6]

ITBS reading comprehension test items are organized around three main process skills: Factual Meaning, Inference and Interpretation, and Analysis and Generalization. **Factual Meaning** involves recall of details and relationships from the text to formulate literal meanings, but little additional knowledge and experience is required to understand the text. **Inference and Interpretation** includes such processes as inferring, deducing, concluding, predicting, applying, determining feelings and motivations of characters, and interpreting non-literal language. This may require the use of textual information to draw conclusions that are implied but not specifically stated. **Analysis and Generalization** requires the reader to use analytical skills and to synthesize information regarding the main idea, topic or theme to determine the author's purpose or viewpoint or to analyze stylistic or structural aspects of a text. According to the published ITBS classification schemes, between 35 and 46 percent of items are classified as Factual Meaning, 32 to 35 percent as Inference and Interpretation, and 22 to 31 percent as Analysis and Generalization.[7]

ISAT questions are designed to align with the Illinois Learning Standards, which were created to guide English/Language Arts curricula in Illinois schools. The ISAT covers five or six content areas (depending on grade level).[8] **Comprehension of Literary Works** includes both fiction and non-fiction passages that are taken from sources such as novels, short stories, and periodicals. **Comprehension of Informational Text** includes only non-fiction passages taken from newspapers and trade journals and covers such topics as scientific or social phenomena. Autobiographies, biographies, personal essays, and speeches may function either as literary or informational text. **Application of Strategies**: **Explicit Ideas** involves identifying important information directly stated in the text. **Application of Strategies: Implicit Ideas** involves analyzing important information in the text to draw logical conclusions. **Vocabulary** involves using contextual clues and other vocabulary skills to understand the key words, phrases, and concepts in literary and informational text. **Word-analysis** involves phonics, word patterns, and other word analysis skills to recognize new words and is included on the third-grade test only. Unlike the ITBS, most ISAT test items are classified into more than one of these areas (and similarly mapped onto more than one learning standard).

Because of the differences in how test items are classified on the two tests (and the fact that each ISAT item is classified into multiple skills categories), it is difficult to make a direct comparison between them. Because of the way the ISAT counts items, it appears that there is more of everything. At the same time, however, the ISAT area for Application Strategies: Explicit Ideas is very similar to the ITBS Factual Meaning category. As noted above, between 35 and 46 percent of ITBS items are in this category. There are somewhat fewer (about 33 percent) on the ISAT, based on the classifications provided on sample tests. As a result, somewhat more of the ISAT items are classified as Application of Strategies: Implicit Ideas than in the two comparable categories on the ITBS. These differences are not large, however.

Both the ITBS and the ISAT intend to include a variety of literary genres. Since the ITBS includes eight passages for each grade level, it shows a greater diversity of genres than the ISAT, which is composed of three to four long passages for each level. The ITBS passages are relatively short, but include at least one poem at each level. The ITBS also tries to cover a wide range of topics appropriate to the target grade level. Similar to the ITBS, sources for ISAT reading passages range from high-interest, grade-appropriate periodicals to newspapers, short stories, and novels.

The ITBS battery has a separate decontextualized dictionary-type vocabulary section and so it includes only a few context-embedded (or context-dependent) vocabulary items on the reading comprehension test. (The vocabulary section of the ITBS is not required in CPS.) In comparison, the ISAT contains more context dependent than application type (usage oriented) items.

Content Differences: Mathematics

We have seen how ITBS and ISAT reading tests differ in terms of their purpose and format. These differences are also reflected in the mathematics test content. The ITBS is oriented toward measuring students' basic mathematics skills and has a much heavier emphasis on computation than the ISAT. According to the *Iowa Tests Interpretive Guide for Teachers and Counselors,* although some mathematics educators believe that "mathematics concepts and problem solving are given too little time relative to computation," computation is still an essential skill.[9] Therefore, the ITBS has one whole subtest on arithmetic computation

Can the Length of a Reading Passage Affect Student Performance?

There are both advantages and disadvantages to a test having a few longer reading passages instead of several short ones. One the one hand, longer passages are more "life like" and more similar to the level at which students must be able to read to succeed academically in the long term. On the other hand, students are presented with less variety and fewer literary genres and there is a greater chance that the selected passages may produce biased responses. That is, some students may find a particular passage uninteresting and not be as motivated to read it carefully all the way through.

and measures "applied computation" on the other two subtests. As a result, fewer items on the ITBS are classified as algebra, geometry, measurement, or data analysis.

Items on the ISAT, on the other hand, are developed from the Illinois Learning Standards and are based on the philosophy that "mathematics is much more than a collection of concepts and skills."[10] The learning standards envision students being able to use basic operations, but they also stress that students should be able to "confront more involved calculations in algebraic and statistical settings."[11] The standards also stress the importance of "connections" across the mathematical disciplines. Mathematical problems are presumed to "connect the concepts of numbers and their operations, measurement, geometry, data, and algebra."[12] Because of this underlying framework, the ISAT places less emphasis on straightforward arithmetic computation. Instead, single items typically require the use of multiple skills from different content areas and computation skills are most often embedded in an algebraic, geometric, statistical, or probabilistic context.

In Table 3, we compare the content areas of the two tests.[13] The table is organized around the three subtests of the ITBS (A. Math Concepts and Estimation, B. Math Problem Solving, and C. Data Interpretation, and Math Computation) and shows corresponding ISAT content classifications. As with reading, it is difficult to make a strict comparison of how much of the different types of content is included in each of the tests. This is primarily because the ISAT "double counts" items based on which areas they cover, whereas ITBS items are counted in only one content area. This difficulty aside, it is clear that the ITBS is much more heavily weighted toward computation, and that the ISAT contains up to three times as many items on algebra, geometry, measurement, and probability and statistics.

The ITBS contains a section called "problem solving" and indeed the test publisher points out the importance of critical thinking skills in mathematics.[14] In comparison, the ISAT has a section called Extended Response, which presents students with a complex problem and requires them to come up with a strategy, arrive at a solution, and explain how and why they took the steps they did. There are two such problems at each grade level. Students' responses on these items are evaluated according to their mathematical knowledge, their strategy, and their ability to explain their reasoning. Many of these problems have more than one correct solution and, in general, they cannot be solved using "standard" algorithms, although some of the eighth-grade sample questions could be done using standard algebraic methods.[15]

Given the ISAT's lesser emphasis on computation and added requirements for student reading, we expected to find a significantly higher correlation between reading and mathematics on the ISAT than on the ITBS. In other words, we thought that the ISAT mathematics score might be more

Table 3. Content/Process Classifications for Items on Mathematics Tests

ITBS	ISAT
A. Number properties/operations Represent, compare, order numbers Use place value Use expanded form Number properties Classify by divisibility Perform operations	**Estimation/Number Sense/ Computation** Understand numbers and their representations
Estimation Standard rounding Order of magnitude Number sense	Estimate using mental mathematics
Algebra Use/interpret operational/ relational symbols Solve equations to model situations Use expressions to model situations Understand/explore patterns	**Algebraic Patterns and Variables** Construct and solve problems using variables Describe/extend geometric/ numeric patterns
	Algebraic Relationships/ Representations Interpret and describe numerical relationships using words, tables, coordinate graphs, symbols, equations, function notations
Geometry Identify, classify, compare geometric figures Apply concepts of area, volume,perimeter[a] Describe geometric properties, patterns, and relationships	**Geometric Concepts** Identify/describe points, lines, angles, 2- and 3-dimensional shapes and their properties, including symmetry, sides, faces, vertices
	Geometric Relationships Identify, describe, classify, and compare figures, including similarity/congruence
Measurement Measure length/distance, time, volume, mass, weight, temperature Use appropriate units Estimate measurements	**Measurement** Measure and compare quantities using appropriate units Estimate measurements Compute area, surface area, volume[a]
Probability and Statistics Apply probability concepts Apply measures of central tendency and variability	**Probability** Determine, describe, and apply the probabilities of events
B. Data Interpretation Read amounts on scales of bar and line graphs, interpret sectors of circle graphs, locate a cell in a table Compare quantities Interpret relationships and trends	**Data Organization and Analysis** Organize, describe, and make predictions from existing data
No direct comparison ↔	**Extended Response** Mathematical knowledge Strategic knowledge Explanation
Problem Solving ↔ Single step Multiple step Approaches and procedures	**Embedded in above content questions**
C. Computation ↔ Add, subtract, multiply, divide using whole numbers, decimals, and fractions	**No direct comparison**

[a] ITBS classifies computing and volume as geometry; ISAT classifies them as measurement.

strongly associated with reading and verbal skills than the ITBS mathematics score. On the ISAT, the correlations between mathematics and reading range from 0.74 to 0.79 among CPS students, depending on grade. They are marginally lower on the ITBS, ranging between 0.73 and 0.75, again depending on grade. If we exclude the computation portion of the ITBS from the total score, the ITBS reading to mathematics correlations increase slightly, to a range of 0.75 to 0.78, nearly identical to the correlations on the ISAT.[16] These findings rule out the proposition that ISAT mathematics scores are more heavily influenced by reading ability, though it is true that computation alone is much less related to reading than are other mathematics skills that require reading ability.[17]

Summary Comparison

Overall, we have the sense that the ISAT "looks and feels" more demanding or difficult than the ITBS. In both reading and mathematics, the ISAT is longer and contains more items. The ISAT contains extended-response questions that require students to communicate their knowledge and skills. The reading passages are much longer on the ISAT than the ITBS. In mathematics, there are no simple computation problems, and fewer single-step "word problems." We do not have the evidence to make a scientific case for this argument (which would require detailed content analysis of all test items), but our *prima*

facie comparison suggests this difference. Going forward, a more rigorous examination of all ISAT items, including their psychometric properties, would make for a stronger comparison between the tests.[18]

Scoring the Tests

The primary reporting score for the ITBS is the Developmental Standard Score (also called a Standard Score or SS). This score is obtained from tables prepared by the test publisher that convert raw scores (number correct) into standard scores. There is a separate table for each level of the test. The Standard Score scale spans all grade levels (kindergarten through eight) and ranges in value from 110 to 350. The national average standard score for third graders is 185 and it is 250 for eighth graders.[19] Scale scores are converted to percentile rankings depending on the time of year the test is administered and the grade level of the student.

ISAT raw scores are also converted to scale scores, ranging in value from 120 to 200. The raw score consists of the number of correct multiple-choice questions plus points from the extended-response questions. On the reading test, extended responses are scored on a 0 to 4 scale. On the mathematics test, they are scored on a 0 to12 scale where each of three subscales is rated from 0 to 4. Each extended-response question "counts" the same as approximately six multiple-choice questions. Thus, the extended-response questions account for about 15 percent of the total score on each test. It should be noted that extended-response questions did not count as part of the total ISAT score until 2001.

Unlike the ITBS, the ISAT scale does not span across test levels and so we cannot measure student progress from one administration of the test to another. This would be difficult anyway, given that the ISAT is not administered to all grades. Whereas a specific score on the ITBS is converted to a percentile ranking to compare a student's score to other students, on the ISAT the scale score results are converted to a category rating that describes a student's level of proficiency with the content on the test. In descending order, these categories are: Exceeds Standards, Meets Standards, Below Standards, and Academic Warning.

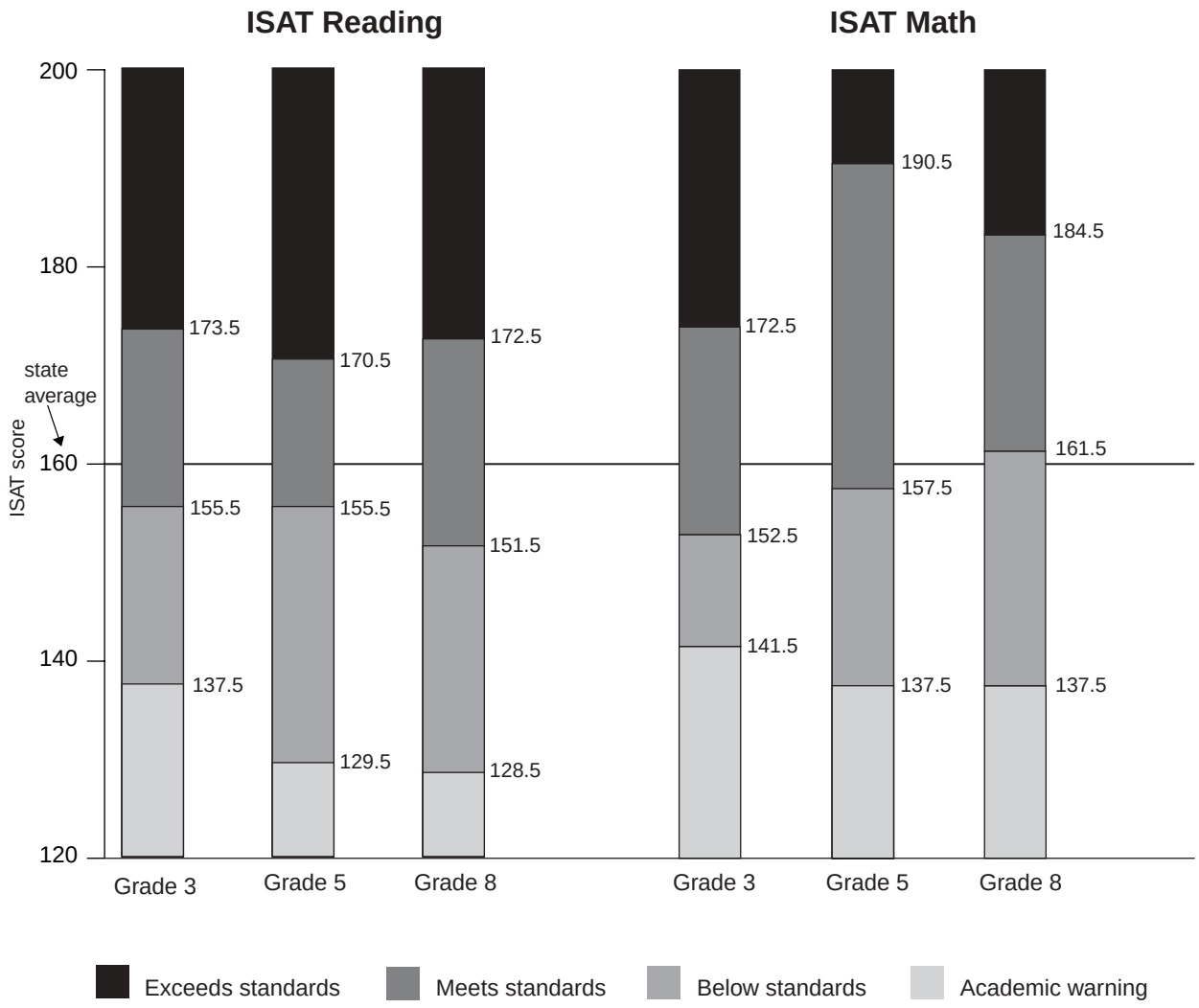
The highest scale scores warrant an Exceeds Standards designation; progressively lower scores result in lower categories. The distinctions between categories, or cut scores, were determined by a two-stage rating process. First, committees of curriculum experts described the expected skill levels of students in each of the four performance categories.[20] Then, teachers and other educators and experts judged each test item. They estimated what percent of students just above each of the three cut scores would answer each item correctly. This process resulted in cut scores on a 120 to 200 scale that demarcates the transitions between categories. (Note that these cut scores were set only once, in 1999. Comparability from year to year in scale score values is maintained by conducting equating studies that link one year's test to the next.)

Figure 1 shows the cut scores between Academic Warning and Below Standards, between Below Standards and Meets Standards, and between Meets Standards and Exceeds Standards for grades three, five, and eight in both reading and mathematics. The range of scores in each category is represented as a length in the stacked bar, moving from Academic Warning at the bottom, to Below Standards, then Meets Standards, and, at the top, Exceeds Standards.

In third-grade reading, scores in the Academic Warning category range from 120 to 137; scores for Below Standards are between 138 and 155; scores for Meets Standards are between 156 and 173; and for Exceeds Standards, scores are between 174 and 200. In third-grade reading, the cut score between Academic Warning and Below Standards is 137.5; the cut score between Below Standards and Meets is 155.5; and between Meets and Exceeds Standards, it is 173.5.[21] Figure 1 also displays the

FIGURE 1

Cut Score from One Performance Category to Another Differs by Grade and Subject

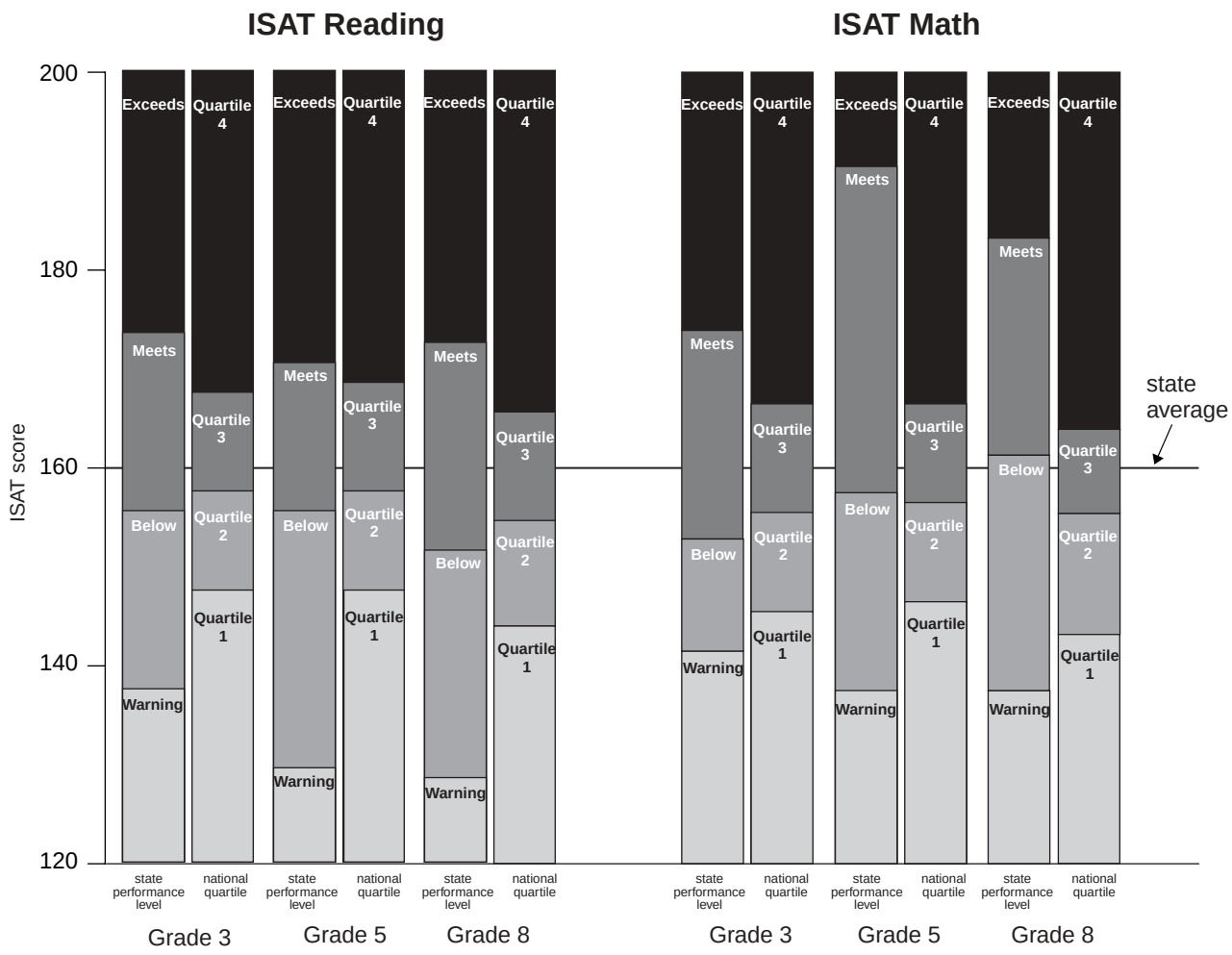


state average score of 160, which was set in the first statewide test administration in 1999.

As seen in the figure, the placement of the cut scores between categories differs depending on both grade and subject. This relates directly to the level of difficulty for a given performance category. Since ISAT results are most frequently reported as the percent of students who meet or exceed state standards in a particular subject, the most important cut score is between Below Standards and Meets because all scores above that point will be counted in the combined category of Meets or Exceeds Standards. In reading in both third and fifth grade, the cut score is 155.5. In eighth grade, it is somewhat lower at 151.5.[22] In mathematics, the cut scores become progressively higher moving up the grades. In other words, a higher score is required to meet state standards. For third grade it is 152.5, for fifth 157.5, and for eighth 161.5. The eighth-grade cut score for meeting or exceeding standards in mathematics is higher than the 1999 state average when the reading and mathematics tests were first administered.

There is also variability in the cut scores that differentiate the Academic Warning category from Below Standards, and the cut scores between Meets

FIGURE 2
Relationship Between Performance Categories and National Quartiles



Standards and Exceeds Standards. For example, only the very lowest performing students in fifth- and eighth-grade reading will warrant an Academic Warning designation, where the cut scores are 129.5 and 128.5, respectively. In contrast, the cut score between Academic Warning and Below Standards is 137.5 for third-grade reading. In mathematics, these cut scores also vary by grade: 141.5 in third, and 137.5 in both fifth and eighth. The difference between a Meets Standards and an Exceeds Standards score also varies considerably by subject and grade. It is most difficult to obtain Exceeds Standards in fifth-grade mathematics, where the cut score is 190.5. Compare this to 172.5 for third-grade mathematics and 170.5 for fifth-grade reading.

This variability should not necessarily be considered arbitrary. As noted, teachers and curriculum experts participated in a rating process to determine the cut scores. It seems that the test makers have decided that students need to know many of the skills on the test in order to meet state standards in mathematics in the higher grades. This might reflect a statewide preference for increasing emphasis on algebraic skills in the upper elementary grades.

As required by state legislation, the ISBE conducted an equating study between the ISAT and the Stanford Achievement Test, Ninth Edition (SAT-9) to compare the state standards to national norms.[23] This requirement was instituted in order

to provide a "check" on the validity of the ISAT and the standards setting process in general. The results of this equating are shown in Figure 2, which compares the four ISAT categories to the four national quartiles (as determined by the SAT-9). There are some general patterns in the relationship between the four state categories and the four national quartiles. For example, Academic Warning is always a subset of the first (lowest) national quartile, and Exceeds is always a subset of the fourth (highest) national quartile. That is to say, the cut score between Academic Warning and Below Standards is always below the 25th national percentile (determined by SAT-9 norming) and the cut score between Meet Standards and Exceeds Standards is always above the 75th national percentile. In reading, the third-grade cut score is about equal to the 50th percentile, it drops lower in fifth grade, and lower yet in eighth. As we have noted, the mathematics cut scores between Below Standards and Meets Standards become progressively higher, from below the 50th percentile in grade three, to about the 50th percentile in grade five, to considerably above the 50th percentile in grade eight. In other words, students need to perform better than the national average (according to SAT-9 norms) and better than the state average in order to achieve a Meet Standards score on eighth-grade mathematics.

Test Equating

ITBS scores are equated across levels and different forms so that scores will be equivalent regardless of the particular level or form a student takes. Such equating permits comparisons over time as students progress from one test level to the next and as the school district administers different forms from year to year. The ISAT is also equated across years, but not across levels.

Equating is an essential feature of standardized tests that enables us to study trend lines over time and to measure students' growth from one year to the next. A recent independent study of the ISAT concluded that equating procedures were rigorous and reliable.[24] Test equating is technically very difficult, however, and in Chicago there have been times when there have been questions about ITBS and Illinois Goals Assessment Program (IGAP, the predecessor to the ISAT) equating. For example, some observers have suggested that certain forms of the ITBS are "harder" or "easier" than others and patterns in test results support this observation.[25] Equating problems can be exacerbated when forms are not directly linked to each other, but are indirectly linked through statistical procedures. An earlier Consortium study also noted this problem.[26] Similarly, another group of researchers hypothesized that declines in state IGAP reading scores were in fact the result of equating problems.[27]

Who is Tested?

Both the ITBS and ISAT seek to test most of the students enrolled in CPS, but their rules for which students must take the test and whose scores are included in public reporting of results differ. On the ITBS, about 94 percent of the total enrollment in the target grades is tested, though only about 74 percent of those tested are included in public reporting.[28] There are two reasons for the 20 percent discrepancy between the number of students who take the ITBS and the number whose scores are publicly reported (those students who take the test but whose scores are not reported are also known as "tested and excluded"). First, some students who receive special education services take the ITBS with accommodations, such as individualized test administration or additional time, as specified in their Individual Education Plans (IEPs). Since many of these accommodations may invalidate the normative interpretation of test results, these students are not included in reporting. Second, a large number of bilingual education students take the test (after two full years in a bilingual program), but their scores are not included in public reporting until after the end of their fourth year in bilingual

education. Untested students include students in special education with severe/profound disabilities and students learning to speak English who have been enrolled in bilingual education classes for fewer than two years.

On the ISAT reading and mathematics tests (given at grades three, five, and eight), fewer students take the test, but scores are reported for all students who take it. That is, there is no "tested but excluded" category as there is on the ITBS. Reported results may be disaggregated into three groups, however: students with an IEP, Section 504 students, and the remaining students—those with neither an IEP nor Section 504 classification.[29] There is a bigger difference between ITBS and ISAT testing for bilingual students, however. Students do not take ISAT until after *three* years in bilingual education, and again, the scores of all students who take the test are reported. More bilingual students take the ITBS than the ISAT, but the scores of more bilingual students are included in the reporting of ISAT results than in the reporting of ITBS results.

Figure 3 shows how many students took these tests in 2002. Most took both tests and their results were publicly reported. This group includes 62 percent of third graders, 75 percent of fifth graders, and 70 percent of eighth graders. Another group of students took both tests, but their ITBS scores were not publicly reported. This group consists of 5 percent of third graders, 10 percent of fifth graders, and 12 percent of eighth graders. This category grows over the grades as more students receive special education services and are thus "tested but excluded" on the ITBS. There are also a substantial number of third graders who do not take the ISAT but do take the ITBS (although their scores are excluded from reporting). This 14 percent are bilingual students who are in their third year of bilingual education. Relatively few students take neither test (3 to 6 percent depending on grade). There is another group of about 3 to 6 percent we call "other." Most of these students have incomplete test data—especially on the ISAT, where there are fewer opportunities for making up a test due to absence or illness (See the Appendix for more information).[30]

Figure 4 shows test inclusion patterns from 1999 to 2002. A few trends are apparent in these graphs. First, as we've noted, more students take the ITBS than the ISAT in all three grades. Second, though fewer students are tested, more are "included" on the ISAT than on the ITBS. Finally, inclusion rates seem to be increasing slightly over time, especially on the ISAT. The notable exception occurs among

FIGURE 3

2002 Test Inclusion Patterns

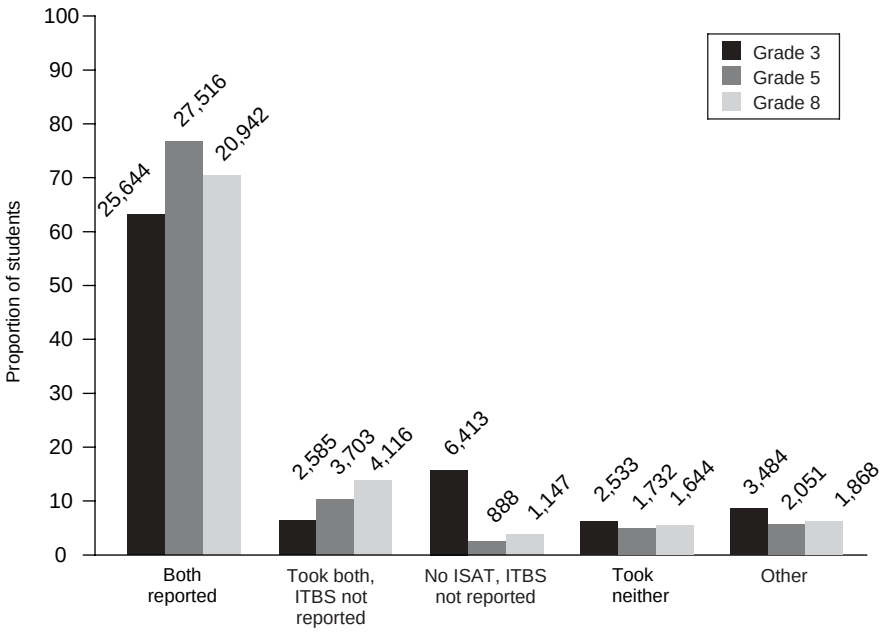
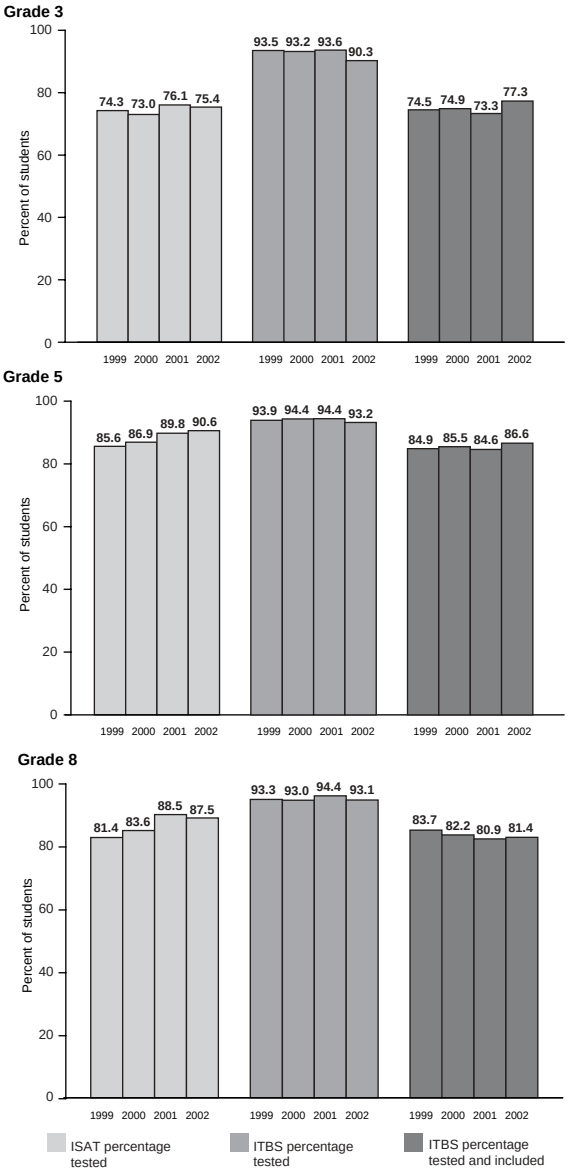


FIGURE 4

Inclusion Trends



eighth graders on the ITBS and this is caused by growing special education enrollments in the upper grades.[31]

Correlations between the ITBS and ISAT

Up to this point, we have compared the ITBS and ISAT in terms of their purposes, content, development, and administration. Given the apparent differences between the two tests, what are our expectations for the relationships between them?

Here we present some empirical evidence from the 2002 CPS spring administrations that links the two tests. We examine the scores of students in grades three, five, and eight in reading and mathematics. For this analysis, and for those in the remainder of this report, our sample is those students who have complete scores on both tests regardless of whether their ITBS scores are included in public reporting.[32]

The two tests are highly correlated. In reading, the correlations for grades three, five, and eight are 0.83, 0.85, and 0.85, respectively. In mathematics, the correlations are slightly higher at 0.86, 0.87, and 0.86. These relationships are comparable to other correlations among standardized achievement tests.[33] When we correct the correlations for unreliability, the correlation coefficients increase to around 0.90.[34]

The correlations suggest that students who perform well on one reading test will perform well on the other and students who perform poorly on one will perform poorly on the other. The same holds true to a somewhat greater degree on the mathematics tests. There are exceptions, of course, but overall, student performance on the two tests is likely to be similar. This does not mean that we can interpret one score in terms of another, however. If a student does well on the ITBS that does not mean we can be sure that he or she would do well on the extended-response questions or the algebra on the ISAT. Similarly, we cannot infer that a student with a high ISAT score knows how to do the timed computations that are required on the ITBS.

Figure 5 compares ITBS and ISAT results for students who took both tests in 2002. The figure shows that students who score in the lower ITBS quartiles are most likely to score in the Academic Warning or Below Standards categories on the ISAT.[35] Similarly, students who score in the higher ITBS quartiles are most likely to score in the Meets Standards or Exceeds Standards categories.

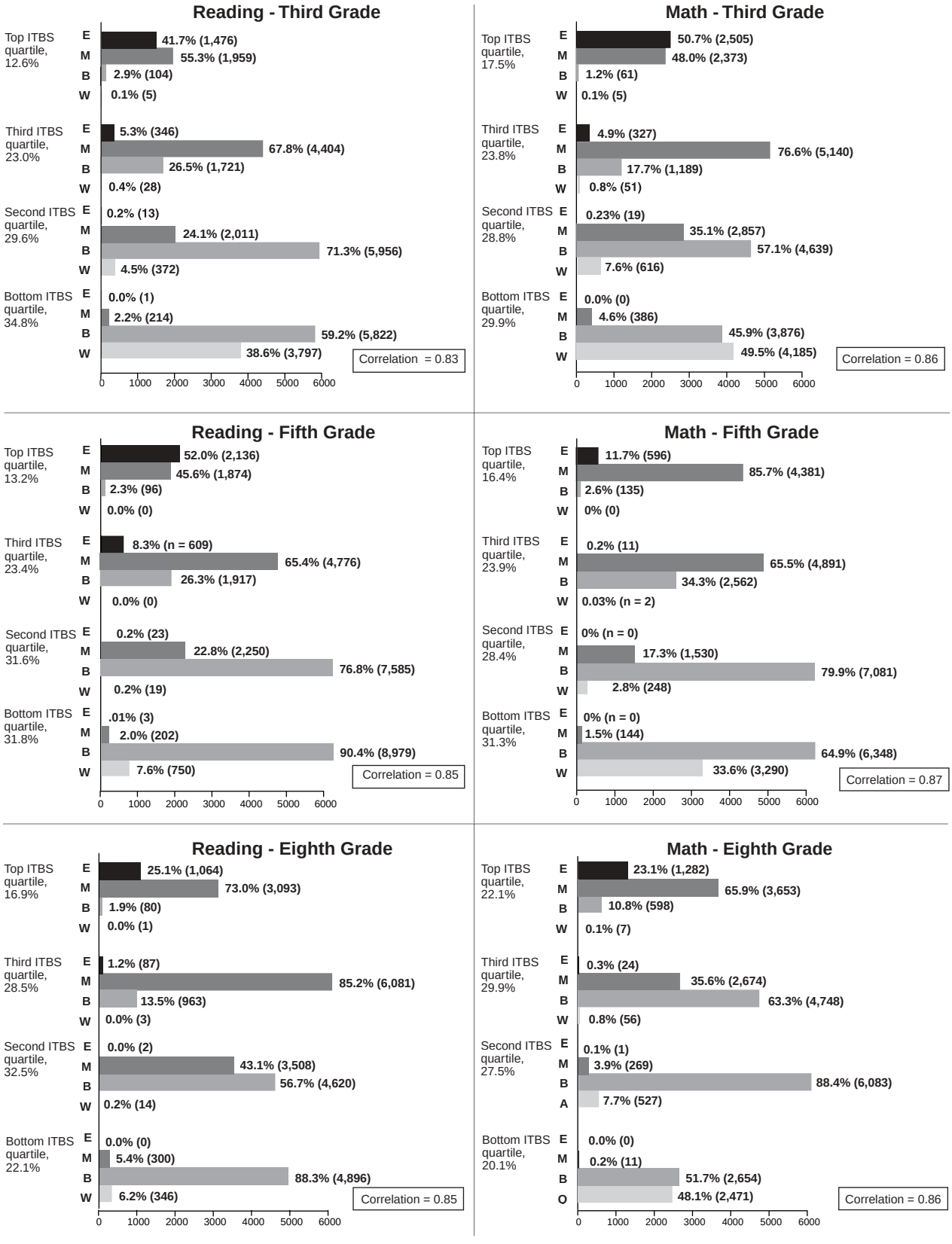
FIGURE 5

ISAT Performance Category Results by ITBS Quartile, 2002

To read these graphs: First, note the ITBS national quartiles along the left side. The graph tells what percent of CPS students scored in each quartile. Then to the right, the bars show what the ISAT scores of those students were.

Exceeds standards | **M**eets standards | **B**elow standards | Academic **W**arning

ISAT Performance Category



In 2002, 35 percent of CPS third-grade students scored in the lowest national quartile on the ITBS reading test. Of those students, the largest portion (59 percent) received a Below Standards score on the ISAT. The next largest group (39 percent) received an Academic Warning.[36] Very few of these students (2 percent) scored a Meets Standards and virtually none met the Exceeded Standards category (though in fact, one student did). Among the 30 percent of students scoring in the second national quartile on the ITBS, the largest group (71 percent) is in the Below Standards category on the ISAT. This pattern holds for students in the remaining quartiles: students with higher scores on the ITBS are likely either to be in the Meet Standards or Exceed Standards categories on the ISAT. It is notable, however, that substantial numbers of students in the third quartile on the ITBS in mathematics do not meet state standards in fifth and eighth grade. This is a result of the high standards in those grades.

In spite of the high correlations between the two tests, one-to-one correspondence does not exist between the ITBS national quartiles and the ISAT performance categories. Students who score in the second quartile of the ITBS are most likely to receive a Below Standards score on the ISAT. Third quartile scorers on the ITBS are most likely to have a Meet Standards score on the ISAT. The correspondence breaks down at the first and fourth quartiles of the ITBS. Many first quartile scorers are in the Below Standards rather than Warning category; more fourth quartile ITBS scorers are in the Meets Standards category rather than Exceeds Standards. As we have seen, this is a function of the standards setting process, which tends to put the Academic Warning designation very low and the Exceeds Standards designation very high.

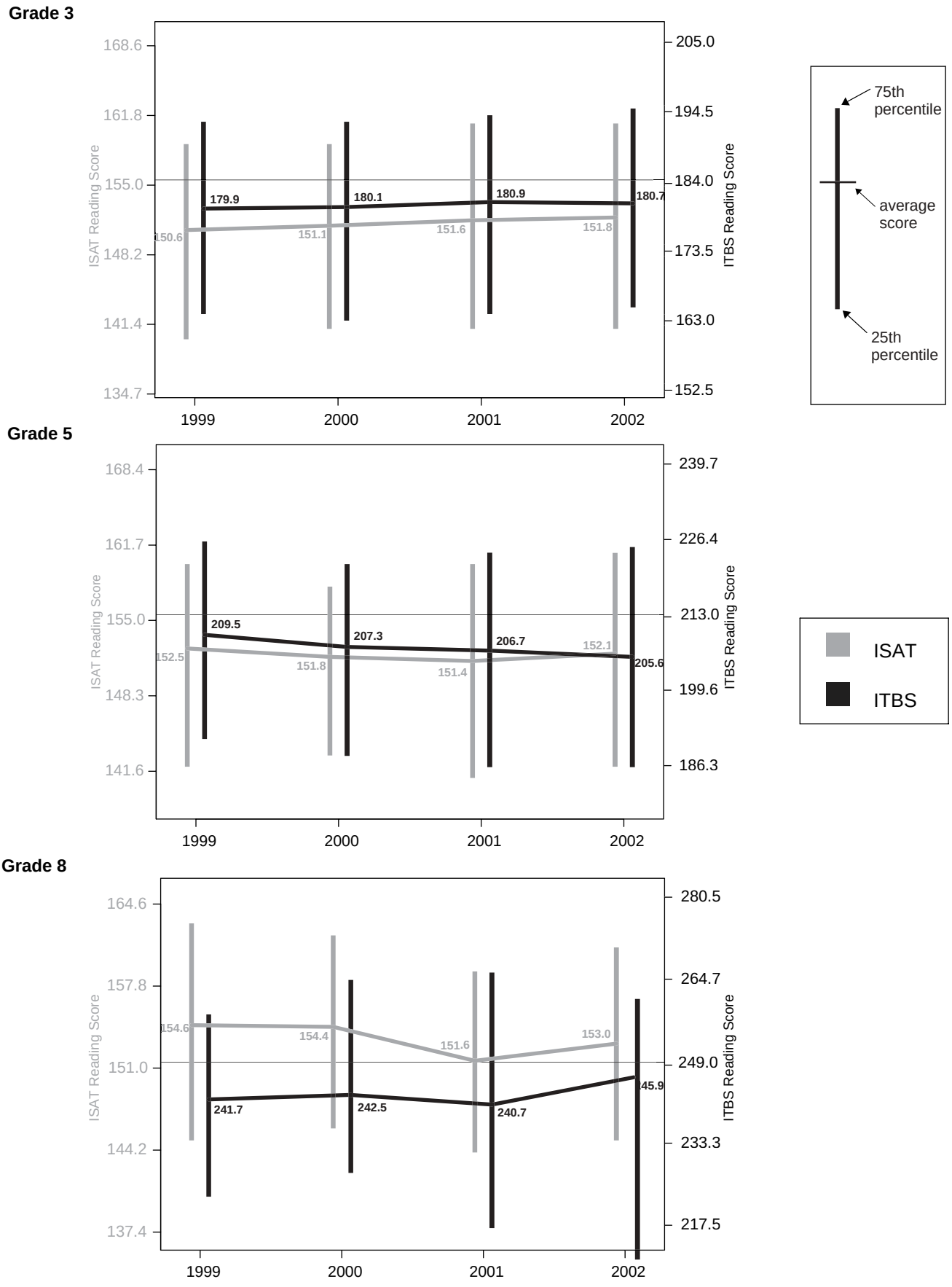
Trends in ITBS and ISAT Results, 1999 to 2002

Frequently, results from the ITBS and ISAT are presented as the percent of students who score at or above a certain cut score. In the case of the ITBS, it is "at or above national norms"; in the case of ISAT, it is the percent of students who "meet or exceed state standards." Though these are useful statistics, they are not fully descriptive of the performance level of the entire population of students under scrutiny. Somewhat more informative are reports of the percent of students in multiple categories that represent a broad range of performance from low to high. With the ITBS, these categories are the four national quartiles and with ISAT they are the four performance categories. The results for these two methods of analysis are presented in the Appendix to this report in Figure 9 and Table 4.

For this report, we are going to discuss a third way of presenting the results. Figures 6 and 7 show the average (mean) scale scores plus the range of scores spanning from the 25th to the 75th percentile (based on the CPS distribution) for grades three, five, and eight in both reading and mathematics. We have argued elsewhere that the arithmetic average is the most useful single statistic for describing group performance because it is influenced by the scores of all students, not just those whose scores are close to the cut scores between categories.[37] For each grade and subject, we present ISAT and ITBS results on the same graph. We performed a type of standardization of the scores to make it possible to compare the test results even though each test has its own scale. First, each test is "centered" on a particular point on its own scale and then these two points are lined up on the graph. For the ISAT, we chose the cut score between Below Standards and Meets Standards, and for the ITBS, we chose the cut score between "at or above

FIGURE 6

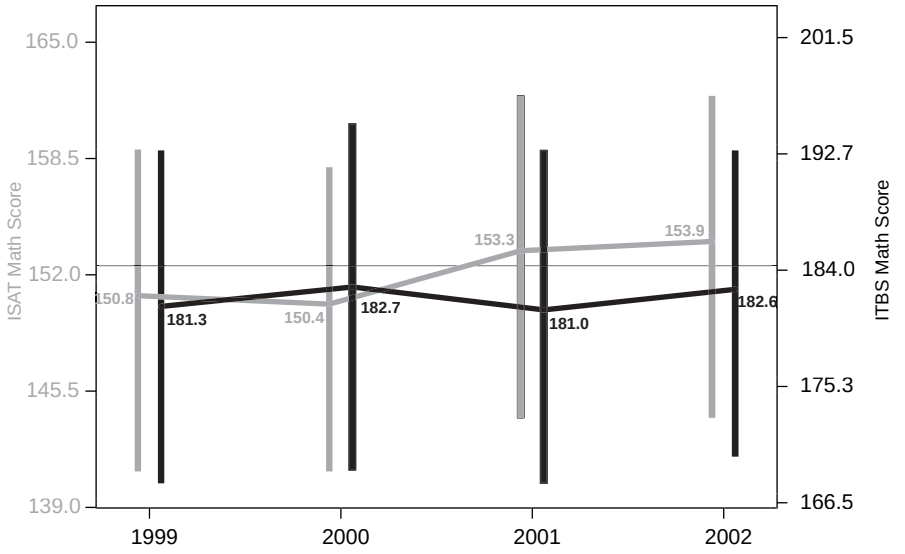
ISAT and ITBS Reading Score Trends

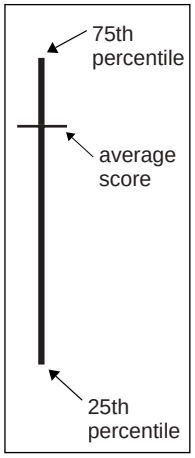


ISAT and ITBS Math Score Trends

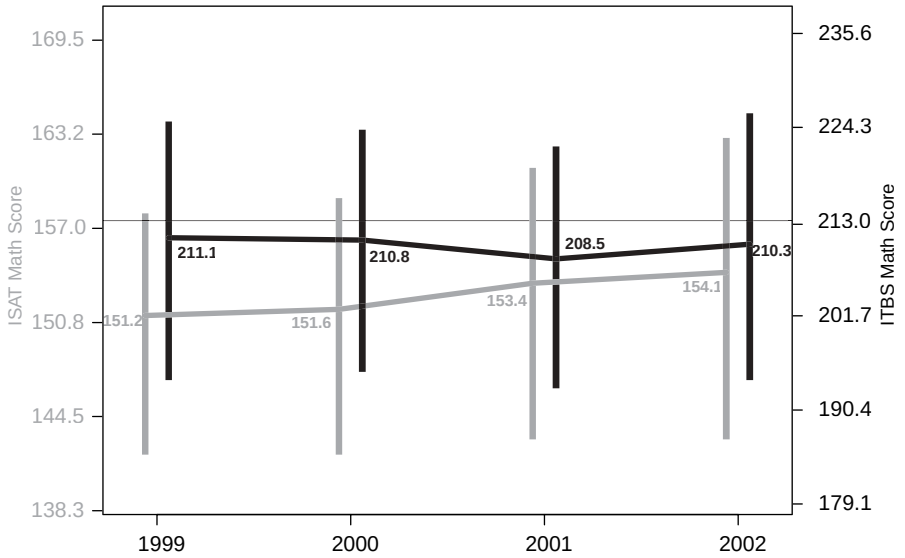
FIGURE 7

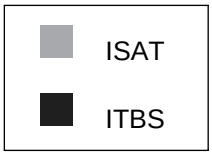
Grade 3



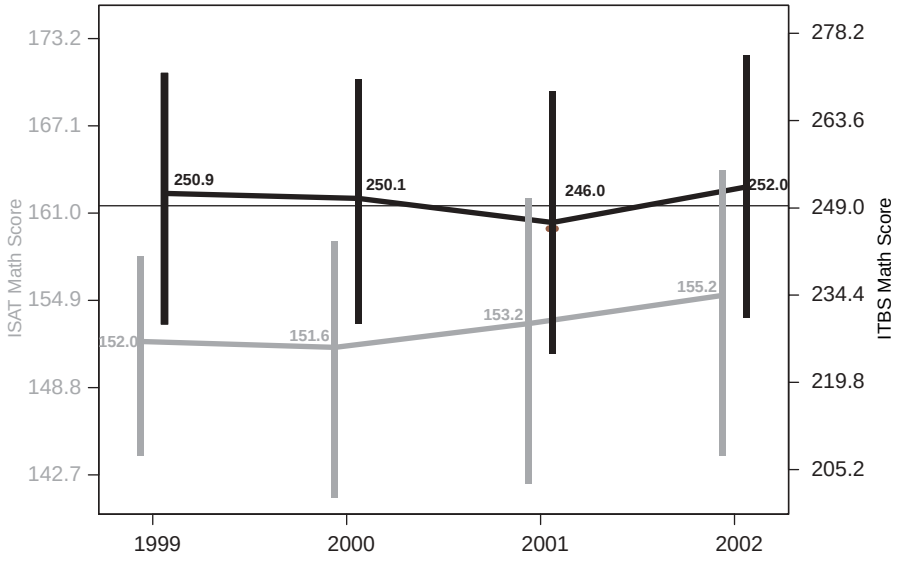


Grade 5





Grade 8



the national average" and "below the national average." So, for third-grade reading, the ISAT cut score of 155.5 is lined up with the ITBS cut score of 184.5.[38] Then, in order to express an equal amount of difference above or below the center, despite the different scales, we calibrated the axes in standard deviation units. That is, one standard deviation below the center on the ISAT is at an equal distance on the vertical from one standard deviation below the center on the ITBS.[39] For third-grade reading, one standard deviation below the center (155.5) is 141.4. The ISAT score of 141.4 is lined up with the ITBS score of 163.0, which is one standard deviation below the cut score of 184.5.

Given this process of standardization, we can compare trends in average CPS performance on the two tests. The graphs display a mixed picture of achievement test score trends in both reading and mathematics for the past four years. Looking at average reading scores across grades and tests, we see an equal number of increases as decreases. In third-grade reading, both ISAT and ITBS scores improved slightly from 1999 to 2002. Fifth-grade reading scores, on the other hand, declined over this period on both tests. In this instance, the ITBS is indicating a significantly greater decline than the ISAT.[40] For eighth-grade reading, the trends on the two tests are running in opposite directions, with the ISAT somewhat down over time, and the ITBS somewhat up.

In mathematics, the results are also mixed, though there are more positive indicators than in reading. In third grade, scores rose (at least slightly) on both tests. In fifth grade, ITBS scores declined and ISAT scores rose. The difference between these opposing trends is statistically significant. Given that the 2001 and 2002 scores on both tests move in the same direction, however, this anomaly may

Difficulties in Linking Different Achievement Tests

Wishing to improve evaluation techniques in national studies, in the early 1970's, the US Department of Education sponsored a large study to equate several standardized achievement test scores onto the same scale.[1] The successful results of this study (which was known as the Anchor study) enabled evaluators to draw conclusions about program effects even though students in different locations took different achievement tests. More recently, as both state and federal governments have discussed or created new testing programs, there have been new attempts to conduct similar studies. These have had markedly different results, however. A 1999 National Research Council report concluded that because there is much greater variability among various achievement tests in terms of scope, content, and format now than there was earlier, it is not feasible (except under special conditions) to report the results of one test on the scale of another.[2] In spite of high correlations, there are enough underlying differences among the tests themselves to invalidate the interpretation of results on one test in terms of another.

The correlation between two tests depends on several factors: the reliability of the tests, the similarity of content and format, and the instructional experiences of the students who take them. When the correlations between tests are high, it is possible to predict performance on one from the other. This does not mean that you can interpret the results of one test in terms of the content of the second.

---

[1] Loret et al. (1974).

[2] Feuer et al. (1999).

not be of much concern. In eighth grade, ITBS scores are flat, but the ISAT trend is up. ISAT mathematics trends are up in all three grades; ITBS trends differ depending on grade.[41] Note how much lower the fifth- and eighth-grade ISAT trends are compared to the ITBS. This is a function of the standards as discussed earlier.

Looking at the trends by grade across tests and subjects, we see all positive results in the third grade, mostly negative in the fifth grade, and mixed in the eighth grade.

The figures also plot scores of relatively high-achieving students (those at the 75th percentile in the Chicago distribution) plus lower-achieving students (those at the 25th percentile) on both the ISAT and ITBS. This additional information allows us to determine whether high- or low-performing students are following the average trends or moving in different directions. In fact, the trends for both groups do follow the average trends discussed above, with very few exceptions. That is, we see the same improvements in third-grade reading on both the ITBS and ISAT for high- and low-performing students as we see in the average scores. Fifth-grade reading looks problematic in all cases, except that ISAT scores for high-performing students may be increasing. Overall, the general patterns are confirmed: increased scores in the third grade in both reading and mathematics on the two tests; negative trends in the fifth grade, with the exception of ISAT mathematics (where there are increases in all three grades); and a mixed picture of performance among eighth graders by subject and by test.

Concluding Observations

In spite of large content and format differences, the ITBS and ISAT behave similarly among CPS students. Their scores are highly correlated and their trends over time are mostly parallel. In the one case where the trends run counter to each other, it looks like they are converging and will be parallel in another year or two.

We have looked only at citywide trends in this data brief, and not trends in individual schools. It is possible, and worth investigating, that the correlations between the ISAT and ITBS may differ from school to school. In that case, it may be very possible in some schools that scores go up on one test while they are going down on the other. Such differences may be the result of differential instructional practices that effect results on one test but not on the other. This does not occur citywide, however.

This examination has also shown how the standards setting process on the ISAT can influence score results. Higher standards will result in fewer students in the Meeting or Exceeding categories, as lower standards will result in more students in those categories. In the eighth grade, ISAT reading scores are high because the cut score is relatively low. On the other hand, eighth-grade ISAT mathematics scores are low because the cut score is set high. This illustrates a major difference between norm-referenced and standards-based approaches. In one case, we accept the national average as our standard of what is "good enough" and have little regard for what the content actually is. In the other case, educators are telling us what we should be expecting of our students independent of normative performance. As NCLB is implemented, this discussion will surely continue.

Appendices

FIGURE 8

Test Inclusion Patterns

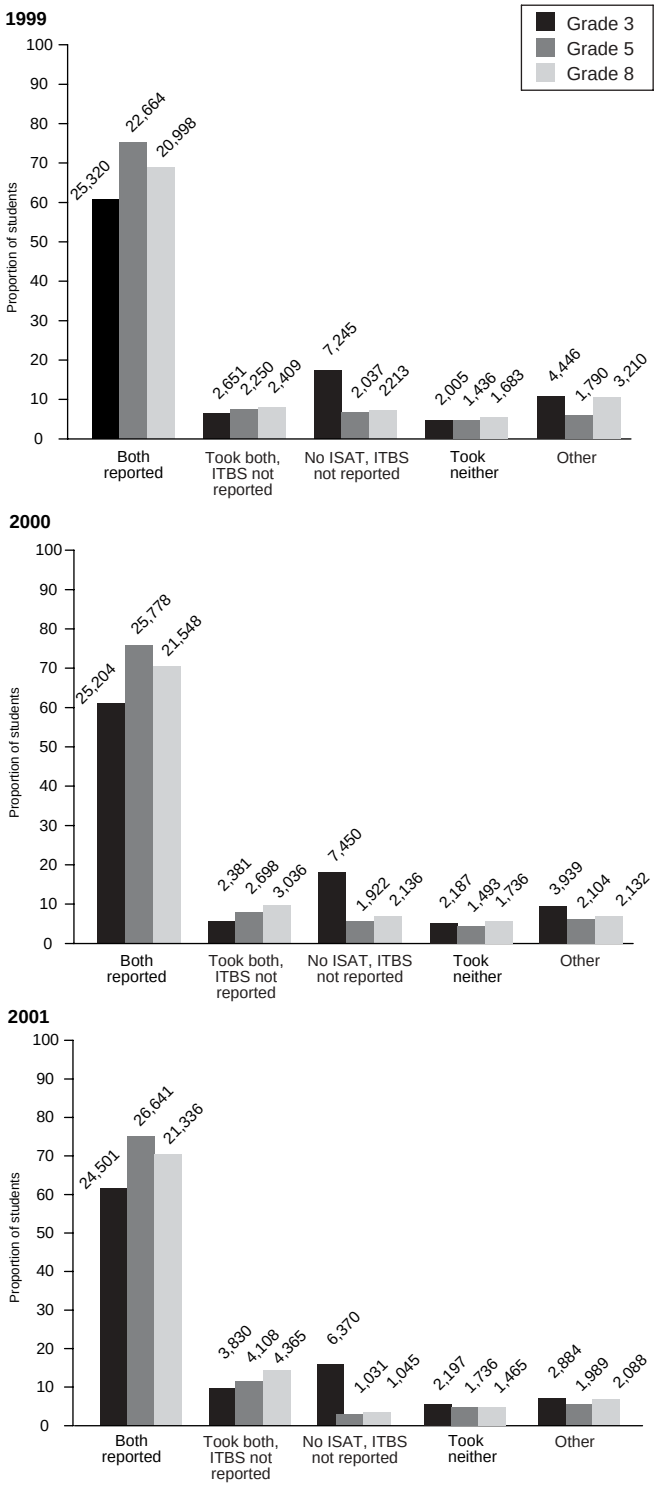


FIGURE 9

ISAT and ITBS Categorical Trends

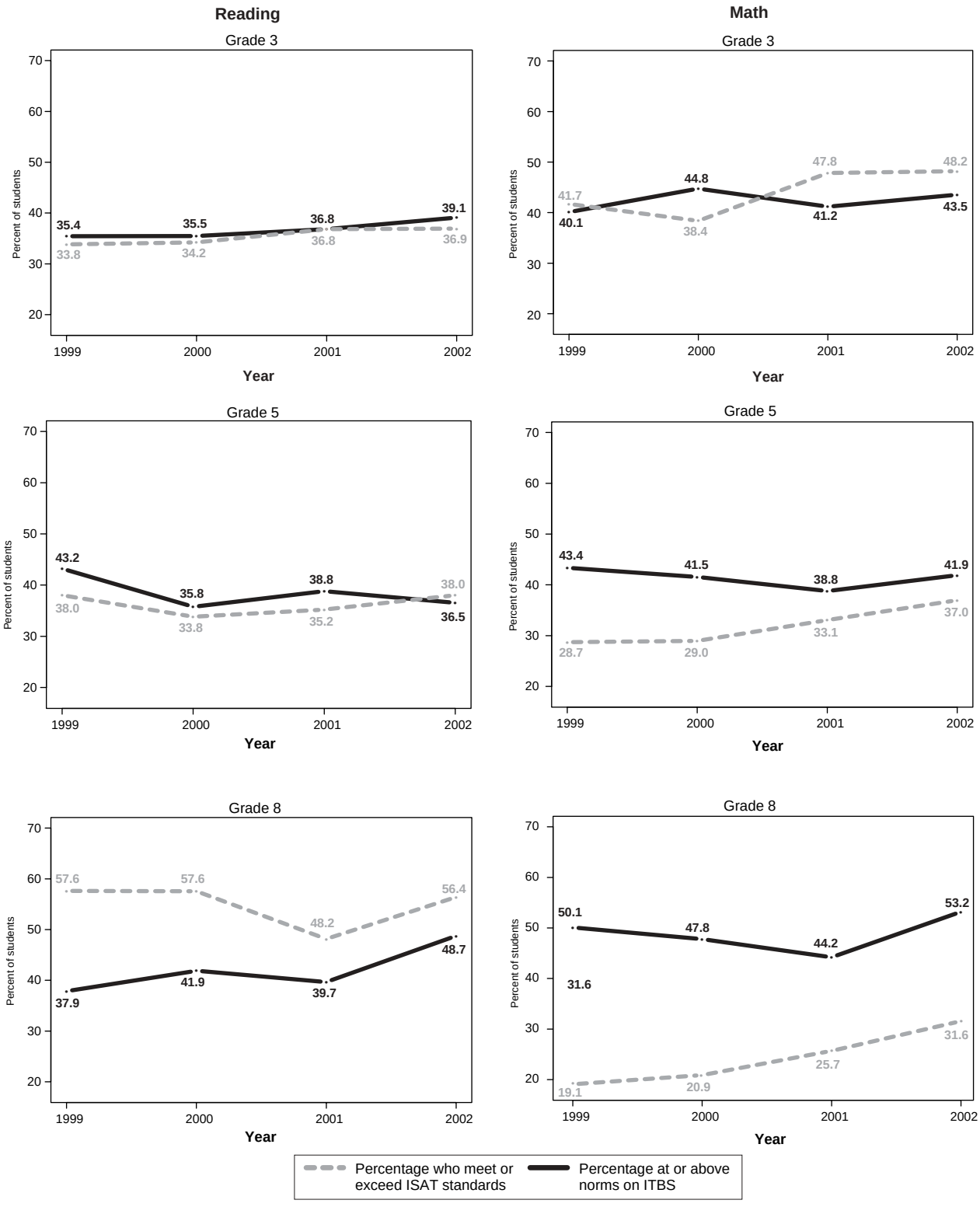


Table 4. ISAT and ITBS Frequencies
1999 - 2002

ISAT

Reading

Grade 3

	1999	2000	2001	2002
Warning	18.6	15.8	16.4	14.9
Below	47.5	50.0	46.8	48.2
Meets	28.7	27.9	30.4	30.4
Exceeds	5.1	6.3	6.5	6.5

Grade 5

	1999	2000	2001	2002
Warning	2.4	0.4	2.1	2.5
Below	59.5	65.7	62.7	59.5
Meets	28.5	27.9	26.0	29.2
Exceeds	9.5	6.0	9.2	8.9

Grade 8

	1999	2000	2001	2002
Warning	1.3	0.4	1.1	1.5
Below	41.1	42.0	50.8	42.1
Meets	48.6	49.7	44.6	51.8
Exceeds	9.0	7.9	3.6	4.6

Math

Grade 3

	1999	2000	2001	2002
Warning	26.2	25.9	19.6	17.2
Below	32.1	35.7	32.5	34.6
Meets	35.3	31.8	38.8	38.1
Exceeds	6.4	6.6	9.0	10.1

Grade 5

	1999	2000	2001	2002
Warning	11.9	12.5	9.6	11.4
Below	59.3	58.5	57.3	51.7
Meets	28.1	27.9	31.7	35.1
Exceeds	0.6	1.0	1.4	1.9

Grade 8

	1999	2000	2001	2002
Warning	9.5	15.1	13.3	12.2
Below	71.4	64.0	61.1	56.2
Meets	17.1	17.7	21.9	26.4
Exceeds	2.0	3.2	3.8	5.2

ITBS

Reading

Grade 3

	1999	2000	2001	2002
First Quartile	34.8	37.2	38.6	34.8
Second Quartile	29.8	27.4	27.9	29.6
Third Quartile	24.8	20.7	19.1	23.0
Fourth Quartile	10.7	14.7	14.4	12.6

Grade 5

	1999	2000	2001	2002
First Quartile	27.3	31.1	33.0	31.8
Second Quartile	33.8	33.2	32.4	31.6
Third Quartile	26.8	25.1	21.1	23.4
Fourth Quartile	12.0	10.6	13.5	13.2

Grade 8

	1999	2000	2001	2002
First Quartile	24.3	24.1	27.9	22.1
Second Quartile	37.8	34.0	32.4	32.5
Third Quartile	25.7	27.6	24.6	28.5
Fourth Quartile	12.2	14.3	15.0	16.9

Math

Grade 3

	1999	2000	2001	2002
First Quartile	34.1	30.8	33.8	29.9
Second Quartile	27.8	26.4	27.0	28.8
Third Quartile	21.2	23.0	21.8	23.8
Fourth Quartile	17.0	19.8	17.4	17.5

Grade 5 Math

	1999	2000	2001	2002
First Quartile	29.3	28.9	32.1	31.3
Second Quartile	29.0	31.2	30.9	28.4
Third Quartile	24.7	23.5	23.0	23.9
Fourth Quartile	17.0	16.4	14.0	16.4

Grade 8

	1999	2000	2001	2002
First Quartile	20.6	20.8	26.5	20.5
Second Quartile	30.4	32.6	30.4	27.5
Third Quartile	29.5	27.9	26.5	29.9
Fourth Quartile	19.5	18.7	16.6	22.1

Endnotes

[1] *No Child Left Behind Act of 2001* (2002). Additional information on this law is available on the Department of Education's website at http://www.ed.gov/offices/OESE/esea/index.html.

[2] Information about the State Assessment Proposal can be found online at http://www.isbe.net/aatf/default.htm.

[3] For more information about CPS's accountability plan, see http://acctplan.cps.k12.il.us/.

[4] This is not meant to be an exhaustive description of the two tests, but rather an overview to provide needed context for understanding the test results. See our sidebar on page 2 for links to other sources that provide in-depth descriptions of the tests. Also note that we have not undertaken any independent analyses of the items in either test, nor have we analyzed the factor structure related to skills and standards sets of items.

[5] Hoover et al. (2001).

[6] Illinois State Board of Education, (2001b).

[7] Hoover et al. (2001).

[8] Illinois State Board of Education (2001b).

[9] Hoover et al. (2001), p. 35.

[10] See, for example, Illinois State Board of Education (2001b), p. 4 and Illinois State Board of Education (2001a), p. 4.

[11] Illinois State Board of Education (2001a).

[12] ibid., p. 8.

[13] The information used to construct the ITBS portion of the table comes from the Hoover et al. (2001), pp. 2 and 146-149. The information used to construct the ISAT portion of Table 3 comes from Illinois State Board of Education (2001a), pp. 26, 40, and 55. Percents are based on the 45 items on the sample test. They are reflective of the whole ISAT to the extent that the content of the samples reflects the content of the complete test.

[14] See Hoover et al. (2001), pp. 15-16.

[15] Sample extended-response questions from 1999 to 2002 are available from the ISBE's website at http://www.isbe.net/assessment/ReleasedMathER.htm.

[16] The correlation between mathematics computation and reading comprehension on the ITBS is between 0.51 and 0.53 depending on the grade. Even with no reading involved, there is a moderate correlation between the two skill areas.

[17] Correlations between mathematics and reading scores are not significantly different among subgroups of students; for example, there is no difference in the correlations for bilingual and non-bilingual students, though one might expect lower correlations for students still learning English.

[18] The Consortium is interested in conducting such an analysis requiring item-level data. We hope to gain access to these data in the future.

[19] The fourth-grade average is 200; fifth grade, 214; sixth grade, 227; and seventh grade, 239.

[20] Information about how the ISAT was created is available from the ISBE's website at http://www.isbe.net/aatf/default.htm.

[21] We defined the cut score as the halfway point between two adjacent performance categories.

[22] Strictly speaking, we should not compare scale scores in one grade to scale scores in another grade because the tests have not been equated across grades. However, scores can be compared in relation to the state average of 160, which was set in 1999.

[23] The equating is described in Illinois State Board of Education (2001b), p. 51.

[24] See Wick (2002).

[25] Form L is thought to be more difficult than Form K. All patterns of test score results support this assertion. See Rosenkranz (2002).

[26] Bryk et al. (1998).

[27] Pearson and Shanahan (1998).

[28] See Rosenkranz (2001).

[29] Section 504 of the Rehabilitation Act of 1973 mandates that no handicapped person may be discriminated against by recipients of federal financial assistance, including primary and secondary schools and universities. It ensures that students with disabilities have access to a free and appropriate education. This often translates into provision of either special education services, or accommodations in a regular classroom. Section 504 differs from the Individuals with Disabilities Education Act in that Section 504 students may not necessarily need special education services and are not required to have an IEP.

[30] The ISAT data files generally contain numerous errors, including invalid student IDs and incorrect designations of race and gender. In contrast, CPS carefully cleans ITBS data files so that they contain relatively few data errors.

[31] See Miller and Gladden (2002), p. 11.

[32] The ISAT scores of all students in our sample have been publicly reported. However, the ITBS results reported here include scores that have not been previously reported.

[33] Feuer et al. (1999), p. 24.

[34] By making this correction, we are estimating the theoretical correlation between the two tests by removing the "noise" created by unreliability in each. In other words, this is what the correlation would be if the two tests were each perfectly reliable.

[35] The first quartile contains percentiles one through 25, the second quartile percentiles 26 through 50, the third quartile percentiles 51 through 75, and the fourth quartile contains percentiles 76 through 99.

[36] There would be more students from the first ITBS quartile in the ISAT Academic Warning category if the cut score between Academic Warning and Below Standards were higher.

[37] See Bryk et al. (1998) and Easton et al. (1998).

[38] This cut score is simply the halfway point between the national average score and the scale score one point lower. In this case, everything above 184.5 is "at or above the national average" and everything below is "below the national average."

[39] The standard deviation is calculated from the CPS distribution of ISAT and ITBS scores in 1999.

[40] $F = 11.45$; $p < 0.01$.

[41] Figure 9 shows the percentage of students above the respective ITBS and ISAT cut scores. The results are similar to results shown in Figures 6 and 7 though they tend to be somewhat more labile from one year to another.

References

Bryk, Anthony S., Yeow Meng Thum, John Q. Easton, and Stuart Luppescu. (1998). *Academic Productivity of Chicago Public Elementary Schools.* Chicago: Consortium on Chicago School Research.

Easton, John Q., Brian Jacob, Stuart Luppescu, and Melissa Roderick. (1998). *Adjusting Citywide ITBS Scores for Student Retention in Grades Three, Six, and Eight.* Chicago: Consortium on Chicago School Research.

Feuer, Michael J., Paul W. Holland, Bert F. Green, Meryl W. Bertenthal, and F. Cadelle Hemphill (Eds.). (1999). *Uncommon Measures, Equivalence and Linkage Among Educational Tests.* Washington, DC: National Academy Press.

Hoover, H.D., S.B. Dunbar, D.A. Frisbie, K.R. Oberley, G.B. Bray, R.J. Naylor, J.C. Lewis, V.L. Ordman, and A.L. Qualls. (2001). *The Iowa Tests Interpretive Guide for Teachers and Counselors.* Itasca, IL: Riverside Publishing.

Illinois State Board of Education. (2001a). *Illinois Standards Achievement Test Sample Mathematics Materials, 2002*. Springfield, IL: Author.

Illinois State Board of Education, Division of Assessment. (2001b). *The Illinois State Assessment 2001 Technical Manual.* Springfield, IL: Author.

Loret, Peter G., Alan Seder, John C. Bianchini, and Carol A. Vale. (1974). *Anchor Test Study.* Washington, DC: US Department of Health, Education, and Welfare.

Miller, Shazia Rafiullah and Robert M. Gladden. (2002). *Changing Special Education Enrollments: Causes and Distribution Among Schools.* Chicago: Consortium on Chicago School Research.

No Child left Behind Act of 2001. US Public Law 107-110. 107th Cong., 1st sess., 8 January 2002.

Pearson, David D. and Timothy Shanahan. (1998) The reading crisis in Illinois. *Illinois Reading Council Journal*. (26) 3:60-67.

Rosenkranz, Todd. (2002). *2001 CPS Test Trend Review: Iowa Tests of Basic Skills.* Chicago: Consortium on Chicago School Research.

Wick, John W. (2002). Independent Assessment of the Technical Characteristics of the Illinois Standards Achievement Test. Evaluation commissioned by the Illinois State Board of Education, Springfield, IL. Available online at http://www.isbe.net/assessment/PDF/WickReport.pdf.

Acknowledgments

Many thanks to Elaine Allensworth, Tony Bryk, Fred Hess, Melissa Roderick, Mark Smylie, and Connie Wise for their thoughtful and helpful comments. Thanks also to Sandra Jennings for her production expertise and to Rose Sweeney for her editing assistance.

Core research activities at the Consortium, including this data brief, are supported by the Joyce Foundation, the John D. and Catherine T. MacArthur Foundation, and the Spencer Foundation.

Consortium on Chicago School Research

Mission

The Consortium on Chicago School Research is an independent federation of Chicago area organizations that conducts research on ways to improve Chicago's public schools and assess the progress of school improvement and reform. Formed in 1990, it is a multipartisan organization that includes faculty from area universities, leadership from the Chicago Public Schools, the Chicago Teachers Union, education advocacy groups, the Illinois State Board of Education, and the North Central Regional Educational Laboratory, as well as other key civic and professional leaders.

The Consortium does not argue a particular policy position. Rather, it believes that good policy is most likely to result from a genuine competition of ideas informed by the best evidence that can be obtained.

Consortium on Chicago School Research
1313 East 60th Street, Chicago, IL 60637
773-702-3364 fax -773-702-2010
www.consortium-chicago.org