Renaissance Place® EDITION

STAR Math®

Computer-Adaptive Math Test and Database

Technical Manual

Assess Student Math Levels in Less Than 12 minutes!

Renaissance Learning
P. O. Box 8036
Wisconsin Rapids, WI 54495-8036

Toll-free phone (for math solutions, such as STAR Math): (866) 492-6284
Toll-free phone (for reading solutions): (866) 846-7323
Toll-free phone (for writing solutions): (866) 220-8811
Local Phone: (715) 424-3636
Fax: (715) 424-4242

Email (general questions): answers@renlearn.com
Email (technical questions): support@renlearn.com
Web site: www.renlearn.com

Copyright Notice

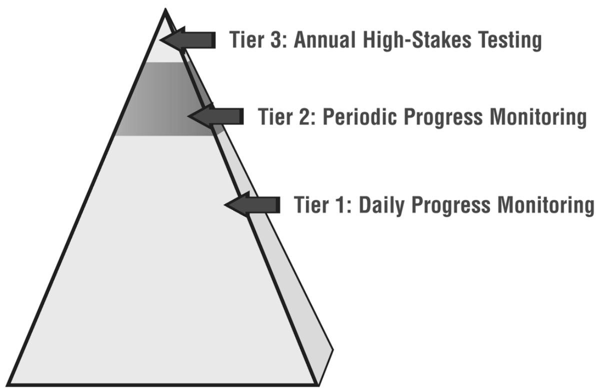
11/03 v3.1

Contents

Chapter 1: Introduction

STAR Math and Learning Information Systems

The STAR Math, Renaissance Place Edition® (RP) computer-adaptive test and database is an achievement-level learning information system (LIS) that allows teachers to assess students' mathematical abilities accurately, in 15 minutes or less. Learning information systems are computer programs that help educators accelerate learning and increase motivation by providing immediate, individualized feedback on student academic tasks and classroom achievement. They operate on the three levels depicted below.



Tier 1: Daily Progress Monitoring

Tier 1 technology provides student-level information on task completion, performance, and time on task. At this most important level, teachers and students receive daily, even hourly feedback. Tier 1 software programs take care of record-keeping tasks, reducing paperwork, and allowing teachers more time to personalize instruction. Students are motivated to correct problems, and their learning accelerates – without adding paperwork for the teacher.

Examples of Tier 1 technology are Renaissance software programs such as Accelerated Reader®, Accelerated Math®, Accelerated Writer®, and Accelerated Vocabulary®, as well as teacher-made assessments.

Tier 2: Periodic Progress Monitoring

Tier 2 periodic progress-monitoring systems help educators place students in material appropriate for their ability and measure growth throughout the year. Used periodically, these assessment tools quickly measure student progress and predict outcomes on mandated high-stakes tests.

Tier 2 software programs include STAR Reading®, STAR Early Literacy®, and STAR Math computer-adaptive tests and StandardsMaster® instant assessments and web-based reporting software.

Tier 3: Annual High-Stakes Testing

National tests and mandated state and district tests, such as the SAT-9 (Stanford Achievement Test, 9th edition), are examples of high-stakes tests. The best way to ensure success on high-stakes tests is through the proper use of Tier 1 and Tier 2 technology.

The formative and diagnostic assessment information provided by Renaissance software is the key to better decision-making, accountability, and higher test scores. At last, this information is easily accessible by all educators and students in the district.

STAR Math Purpose

As a periodic progress monitoring LIS, STAR Math software serves two primary purposes. First, it provides educators with quick and accurate estimates of students' instructional math levels relative to national norms. Second, it provides the means for tracking growth in a consistent manner over long time periods for all students. This is especially helpful to school- and district-level administrators.

While the STAR Math test provides accurate normed data like traditional norm-referenced tests, it is not intended to be used as a "high-stakes" test. Generally, states are required to use high-stakes tests to document growth, adequate yearly progress, and mastery of state standards. These high-stakes tests are also used to report end-of-period performance to parents and administrators or to determine eligibility for promotion or placement. STAR Math is not intended for these purposes. Rather, because of the high correlation between the STAR Math test and high-stakes instruments, classroom teachers can use STAR Math scores to fine-tune instruction while there is still time to improve performance before the regular testing cycle. At the same time, school- and district-level administrators can use STAR Math to predict performance on high-stakes tests. Furthermore, STAR Math results can easily be disaggregated to identify and address the needs of various groups of students.

STAR Math's unique powers of flexibility and repeatability provide specific advantages for various groups:

- For students, STAR Math software provides a challenging, interactive, and brief test that builds confidence in their math ability.
- For teachers, STAR Math software facilitates individualized instruction by identifying students' current developmental levels and areas for growth.
- For principals, STAR Math software provides regular, accurate reports on performance at the class, grade, and building, and district level, as well as year-to-year comparisons.
- For district administrators and assessment specialists, STAR Math software furnishes a wealth of reliable and timely data on math growth at each school and district-wide. It also provides a valid basis for comparing data across schools, grades, and special student populations.

This manual documents the suitability of the STAR Math learning information system for these purposes and presents evidence of its reliability, validity, and merits as a psychometric instrument.

Design of STAR Math

One of the fundamental decisions when designing STAR Math involved the choice of how to administer the test. Because of the numerous advantages offered by computer-administered tests, it was decided to develop STAR Math as a computer software product.

The primary advantage of using computer software to administer the STAR Math test is the ability to tailor each student's test based on his or her specific responses to previous items. Paper-and-pencil tests are obviously far different from this: every student must respond to the same items in the same sequence. Using computer-adaptive procedures, however, it is possible for students to be tested using items that appropriately match their current level of proficiency. Adaptive Branching™, the item selection procedure used in the STAR Math test, effectively customizes every test to the student's current achievement level.

Adaptive Branching offers significant advantages in terms of test reliability, testing time, and student motivation. First, reliability improves over paper-and-pencil tests because the test difficulty matches each individual's performance level; students do not have to fit a "one test fits all" model. With a computer-adaptive test, most of the test items to which students respond are at levels of difficulty that closely match their achievement levels. Testing time decreases because, unlike in paper-and-pencil tests, students need not be exposed to a broad range of material, some of which is inappropriate because it is either too easy for high achievers or too difficult for those with low current levels of performance. Finally, computer-adaptive assessments improve student motivation simply because of the aforementioned issues: test time is minimized and test content is neither too difficult nor too easy. Not surprisingly, most students enjoy taking STAR Math tests, and many report that it increases their confidence in math.

Another fundamental STAR Math design decision involved the format of the test items. The items had to be easily administered and objectively scored by a computer and also provide the breadth of construct coverage necessary for an assessment of math achievement. The traditional four-item multiple-choice format was chosen, based on considerations of efficiency of assessment, objectivity, and simplicity of scoring.

The final fundamental design decision involved determining the organization of the content in STAR Math. Because of the great amount of overlap in content in the math construct, it is difficult to create distinct categories or "strands" for a mathematics achievement instrument. After reviewing the STAR Math test's content, curricular materials, and similar math achievement instruments, the following eight strands were identified and included in STAR Math: Numeration Concepts, Computation Processes, Word Problems, Estimation, Data Analysis and Statistics, Geometry, Measurement, and Algebra.

In versions 2.x and 3.x, the STAR Math test is further divided into two parts. The first part of the test, the first sixteen items, includes items only from the Numeration Concepts and the Computation Processes strands. The first eight test items (items 1-8) are from the Numeration Concepts strand, and the following eight test items (items 9-16) are from the Computation Processes strand. The second part of the test, or the final eight items, includes items from all of the remaining strands. Hence, items 17 through 24 are drawn from the following six strands: Word Problems, Estimation, Data Analysis and Statistics, Geometry, Measurement, and Algebra. The specific makeup of the strands used in the final eight items depends on the student's grade level. For example, a student in grade 1 will not receive items from the Estimation strand, but items from this strand could be administered to a student in grade 12.

The decision to weight the test heavily toward Numeration Concepts and Computation Processes resulted from the fact that these strands are fundamental to all others, and they include the content about which teachers desire the most information. Although this approach emphasizes the two strands in the first part of the test, it provides adequate content balance to assure valid assessment. Additionally, factor analysis of the various content strands supports the fundamental unidimensionality of the construct being measured in the STAR Math 2.x/3.x test; therefore, splitting the test in this way does not impact the measurement validity.

Each STAR Math item was developed in association with a very specific content objective (described in Chapter 2, which begins on page 13). In addition, the calibration trials included items that were expressed differently in textbooks and other reference materials, and only the item formats that provided the best psychometric properties were retained in the final item bank. For example, many questions were crafted both with and without graphics supporting the text of the question. For items containing text in either the question stem or the response choices, great care was taken to keep the text simple and the reading level as low as practical. This is particularly important with computer-adaptive testing because high-performing, lower-grade students may receive higher grade-level questions.

In an attempt to minimize the administration of inappropriate items to students, each item in the item bank is assigned a curricular placement value corresponding to the lowest grade where instruction for this content would occur. During testing, students receive items with a maximum curricular placement value of three grades higher than their current grade level. Although this constraint does not limit the attainable scores in any way, since very difficult items still exist in the item bank within these constraints, it does help to minimize presentation of items for which the student has not yet had any formal instruction.

Improvements to the STAR Math Test in Versions 2.x and 3.x

Since its introduction in 1998, the STAR Math test has undergone a process of continuous research and improvement. Version 2.0 was an entirely new test, with new content and several technical innovations. Version 3.0 RP is an adaptation of version 2.0 designed specifically for use over a computer network, using a web browser. However, the content in STAR Math version 3.0 is identical to the content in STAR Math version 2.x. The following improvements introduced in version 2.0 continue to apply to version 3.0 RP.

- The item bank has been expanded by 38%, from 1,434 items to 1,974 items.
- The content of the item bank has been expanded as well. The item bank now covers 214 objectives, compared to 176 in STAR Math 1.x. Many of the new objectives cover topics in high school algebra, resulting in an improvement in STAR Math's usefulness for assessing college-bound high school students. Other new objectives cover simpler math topics to accommodate the addition of grades 1 and 2 to the STAR Math product.
- The test specifications have been changed to limit the number of items measuring a single objective that may be administered. This ensures diversity in terms of content objectives and provides a more balanced assessment of the math construct.
- Content balancing specifications, grounded in curricula, have been implemented. This ensures that every test will include items assessing student proficiency in a variety of math content areas.
- The distribution of items among Numeration Concepts, Computation Processes, and other applications (all other STAR Math strands) has been changed. In STAR Math 2.x and 3.x, one third of the items in each test will come from each of those three broad areas.
- The difficulty level of the test has been eased to enhance student motivation and minimize student frustration. The STAR Math 2.x/3.x adaptive brancher selects items that each student can answer correctly about 75% of the time. In STAR Math 1.x, the adaptive brancher selected items that each student could answer correctly about 50% of the time. This modification in STAR Math 2.x/3.x results in a testing session with items that are neither too hard nor too easy.

- New norms have been developed to provide the most accurate and up-to-date scores possible.
- The Diagnostic Report has undergone major changes to provide educators with detailed information about each student's current math achievement.
- A new Accelerated Math Library Report has been created that provides educators with a simple method for placing their students in the appropriate Accelerated Math library after a STAR Math test.

Improvements Specific to STAR Math Version 3.x RP

In version 3.x RP, for the first time the STAR Math system is designed for use in a client/server environment, in which all database management functions are controlled from a network server, while tests are administered on computer workstations on the network. All management and test administration functions are controlled using an online management system (Renaissance Place®) which is accessed by means of a web browser from any workstation on the network.

The client/server environment makes a number of new features possible:

- It makes it possible for multiple schools to share a central database, such as a district-level database. Records of students transferring between schools within the district will be maintained in the database; the only information that needs revision following a transfer is the student's school and class assignments.
- The same database that contains STAR Math data can contain data on other STAR tests, including STAR Early Literacy and STAR Reading. Renaissance Place is a powerful, online information management program that allows you to manage all your district, school, personnel, parent, and student data in one place. Changes made to district, school, teacher, parent, and student data for any of these programs, as well as other Renaissance Place software, are reflected in every other Renaissance Place program that shares the central database.
- Multiple levels of access are available, from the test administrator within a school or classroom, to teachers, principals, district administrators, and even parents.
- Renaissance Place takes reporting to a new level. Not only can you generate reports from the student level all the way up to the district level, but also the system allows reports to be limited to specific groups, subgroups, and combinations of subgroups.

  This supports "disaggregated" reporting; for example, a report might be specific to students eligible for free or reduced lunch, to English language learners, or to students who fit into both categories.

  It also supports compiling reports by teacher, by class, by school, by grade within a school, and by many other characteristics such as a specific date range.

- Since Renaissance Place is accessed through a web browser, teachers (and administrators) will be able to access the program from home – provided the district or school gives them that access.
- Whenever a client program (such as the STAR Math 3.0 RP Student program) is started on a workstation, it will automatically check with the server for updates to the program. If there is an update, the client program will automatically be updated without need for any action by the user.

Test Security

STAR Math software includes a variety of features intended to provide adequate security to protect the content of the test and to maintain the confidentiality of the test results.

Split Application Model

STAR Math software separates the various functions into two distinct programs: a Student program and the browser-based management program. The Student program has very few functions; it simply enables teachers to administer the test to students without providing access to confidential information. The browser-based management program, on the other hand, is the administrative portion of the software. It allows teachers and administrators to manage student and class information, customize the program, register students for testing, and create informative reports about student test performance.

Individualized Tests

Using Adaptive Branching, every STAR Math test consists of items chosen from a large number of items of similar difficulty based on the student's estimated ability. Because each test is individually assembled based on the student's past and present performance, identical sequences of items are rare. This feature, while motivated chiefly by psychometric considerations, contributes to test security by limiting the impact of item exposure.

Data Encryption

A major defense against unauthorized access to test content and student test scores is data encryption. All of the items and export files are encrypted. Without the appropriate decryption code, it is practically impossible to read the STAR Math data or access or change it with other software.

Access Levels and Capabilities

Each user's level of access to a Renaissance Place program (both the browser-based portion and the client portion) depends on the primary position assigned to that user and the capabilities the user has been granted in Renaissance Place. Each primary position is part of a user group. There are seven user groups: district administrator, district staff, school administrator, school staff, teacher, parent, and student. By default, each user group is granted a specific set of capabilities. Each capability corresponds to one or more tasks that can be performed in the program. The capabilities in these sets can be changed; capabilities can also be granted or removed on an individual level. Since users can be assigned to the district and/or one or more schools (and be assigned different primary positions at the different locations), and since the capabilities granted to a user can be customized, there are many, varied levels of access an individual user can have.

The security of the STAR Math data is also protected by each person's user name (which must be unique) and password. User names and passwords identify users and the program only allows them access to the data and features that they are allowed based on their primary position and the capabilities that they have been granted. Every person who logs in to Renaissance Place (teacher, school administrator, or district administrator) must enter a user name and password before he or she can access the data and create reports. Without an appropriate user name and password, teachers cannot use the STAR Math programs.

Test Monitoring/Password Entry

Test monitoring is another useful STAR Math security feature. The test-monitoring preferences are implemented using the Testing Password preference and specify whether or not students and test monitors must enter their individual passwords at the start of a test. Because this ensures that students cannot take tests using another student's name, passwords safeguard student data.

Final Caveat

While STAR Math software can do much to provide specific measures of test security, the most important line of defense against unauthorized access or misuse of the program is user responsibility. Teachers and test monitors need to be careful not to leave the program running unattended and to monitor all testing to prevent students from cheating, copying down questions and answers, or performing "print screens" during a test session. They should also ensure that scratch paper used in the testing process is gathered and discarded after each testing session. Taking these simple precautionary steps will help maintain STAR Math's security and the quality and validity of its scores.

Test Administration Procedures

STAR Math 3.x uses the norms developed for STAR Math 2.0. In order to ensure consistency and comparability of test results to the STAR Math 2.0 norms, teachers administering a STAR Math 3.x test should follow the recommended administration procedures. These same procedures were used by the norming participants. It is also a good idea to make sure that the testing environment is as free from distractions for the student as possible.

During STAR Math 2.0 norming, the program was modified so that teachers could not deactivate the proctoring (test-monitoring) options. This was necessary to ensure that the norming data gathered were as reliable as possible. During norming, test monitors had responsibility for test security and were required to provide access to the test for each student. In the final version of the software, teachers can turn off the requirement for test monitoring using the Testing Password preference, but it is not recommended that they do so.

Also during STAR Math 2.0 norming, all of the participants received the same set of test instructions contained in the *Pretest Instructions* included with the STAR Math 3.x product. These instructions describe the standard test orientation procedures that teachers should follow to prepare their students for the STAR Math 3.x test. These instructions are intended for use with students of all ages and have been successfully field-tested with students ranging from grade 1 to grade 12. It is important to use these same instructions with all students prior to STAR Math 3.x testing. While the *Pretest Instructions* should be used prior to each student's first STAR Math test, it is not necessary to administer them prior to a student's second or subsequent tests.

Test Interface

The STAR Math test interface was designed to be both simple and effective. For purposes of standardization, the test limits input to the letter keys (<A>, <B>, <C>, and <D>) on the standard keyboard. Field testing showed that the variability of student proficiency with the mouse and the variance in mouse sensitivity were too great for it to be used as a standard input device. When completing computerized, standardized achievement tests, computer-literate students should have no advantage over those with limited computer skills. Because students have a nearly equal footing by limiting input to only four letter keys and the <Enter> key (or the <return> key on Macintosh computers), these keys were selected for use in the STAR Math product.

Practice Session

The practice session before the STAR Math test allows students to become comfortable with the test interface and to make sure that they know how to operate the software properly. Students can pass the practice session and proceed to the actual STAR Math test by answering two out of the three practice questions correctly. If a student does not do this, the program presents three more questions, and the student can pass the practice session by answering two of those three questions correctly. If the student does not pass after the second attempt, the student will not proceed to the actual STAR Math test.

Even students with low math and reading skills should be able to answer the practice questions correctly. However, the Student program will halt the testing session and will tell the student to ask the teacher for help if the student does not pass after the second attempt.

Students may experience difficulty with the practice questions for a variety of reasons. The student may not understand math even at the most basic level or may be confused by the "not given" response option presented in some of the practice questions. Alternatively, the student may need help using the keyboard. If this is the case, the teacher (or monitor) should help the student through the practice session during the student's next STAR Math test. If a student still struggles with the practice questions with teacher assistance, he or she may not yet be ready to complete a STAR Math test.

Adaptive Item Selection

STAR Math's item selection branching algorithm uses a proprietary approach somewhat more complex than the simple Rasch Maximum Information IRT model. The approach used in the STAR Math test was designed to yield reliable test results by adjusting item difficulty to the responses of the individual being tested while striving to minimize test length and student frustration.

As an added measure to minimize student frustration, the first administration of the test begins with items that have a difficulty level substantially below what a typical student at a given grade can handle, usually one or two grades below grade level.

Teachers can override the use of grade placement for determining starting difficulty by entering the current level of mathematics instruction for the student using the MIL (Math Instruction Level). When an MIL is provided, the program uses that value to raise or lower the starting difficulty of the first test. On the second and subsequent administrations, the test begins about one grade lower than the ability last demonstrated within six months.

Once the testing session is underway, STAR Math software administers 24 items of varying difficulty, adapting the difficulty level of the items dynamically according to the student's responses. The average length of time required to complete a STAR Math 2.0 test is between 11 and 12 minutes, with a standard deviation of about four minutes. Since the content and Adaptive Branching in STAR Math 3.x are identical to that in STAR Math 2.0, test administration time for the two products should be similar. It should be noted that unlike traditional tests, the time required for completion increases with ability. For example, students performing at and above the 90th percentile will on average require about 13 minutes to complete the test, while students performing at or below the 10th percentile require only 10 minutes.

Test Repetition

Students can take a STAR Math test up to five times per year without concern for previous exposure to the items. The STAR Math 3.x item bank contains more than 1,900 items created from eight different content strands. Because the STAR Math software keeps track of the specific items presented to each student from test session to test session, it does not present the same item more than once in any six-month period. By doing so, the software keeps item reuse to a minimum. In addition, if a student is progressing in mathematics development throughout the year and from year to year, item exposure should not be an issue at all. More information on the content of the STAR Math 3.x item bank is available in Chapter 2 (page 13).

Item Time Limits

The STAR Math test has a fixed three-minute time limit for individual test items and a fixed ninety-second time limit for practice items. A fixed time limit was chosen to avoid the complexity and confusion associated with a variable time-out period. Three minutes was chosen on the basis of calibration and norming timing data and general content testing experience.

When a student has only 15 seconds remaining for a given item, a picture of a clock appears in the top right-hand corner of the screen, indicating that he or she should make a final selection and move on. Items that time out are counted as incorrect responses unless the student has the correct answer selected and does not press <Enter> (or <return>) before the item times out. If the correct answer is selected at that time, the item will be counted as a correct response.

The items were crafted with one minute as the maximum amount of time that a student who knew how to do the mathematics would require to complete the solution and respond. During STAR Math 2.0 norming, the mean item response time was 27 seconds with a standard deviation of 25 seconds. The median was 19 seconds, and nearly all (99.7%) item responses were made within the three-minute time limit. Mean and median response times were similar at all grades. Although the incidence of maximum time limits was somewhat higher at the lowest three grades than in other grades, fewer than half of one percent of item responses reached the time limit. This was true even for first-grade students. This suggests that the time limits used for STAR Math 3.x allow ample time for nearly all students to complete the questions.

Chapter 2: Content and Test Design

Content of the STAR Math test evolved through three stages of development. The first stage of development involved specifying the curriculum content to be reflected in the test. Because rules for writing the items influenced the exact ways in which this content finally appeared in the test, these rules may be considered part of this first stage of development. The following section describes these rules. In the second stage, items were empirically tested in a calibration research program, and items most suited to the test model were retained. The third stage occurs dynamically as each student completes a STAR Math test. The content of each STAR Math test depends on the selection of items for that individual student according to the computer-adaptive testing mode.

Content Specification

STAR Math test content was intended to reflect the objectives commonly taught in the mathematics curriculum of contemporary schools (primarily in the United States). Four major sources helped to define this curriculum content. First, an extensive review of content covered by leading mathematics textbook series was conducted. Second, state curriculum guides or lists of objectives were reviewed. Third, the *Principles and Standards for School Mathematics* of the National Council of Teachers of Mathematics (NCTM) was employed. Finally, content specifications from the National Assessment of Educational Progress (NAEP) and the Trends in International Mathematics and Science Study (TIMSS) were consulted. There is reasonable, although not universal, agreement among these sources about the content of mathematics curricula.

The final STAR Math content specifications were intended to cover the objectives most frequently found in these four sources. In the end, the STAR Math content was organized into eight strands. Two hundred fourteen objectives were then created within these eight strands. Appendix A (page 106) lists the specific objectives in each strand.

Numeration Concepts

The Numeration Concepts strand encompasses 43 objectives, making it the strand with the largest number of objectives. This strand concentrates on conceptual development of the decimal number system. At the lowest levels, it covers cardinal and ordinal numbers through ten (the ones). The strand then proceeds to treatment of the decades (tens), hundreds, thousands, and then larger numbers such as hundred thousands and millions, all in the whole-number realm. At each of these levels of the number system, specific objectives relate to place value identification, number-numeral correspondence, and expanded notation. Following treatment of the whole numbers, the Numeration Concepts strand moves to fractions and decimals. Coverage includes representation of fractions and decimals on the number line, conversions between fractions with different denominators and between fractions and decimals, number-numeral correspondence for decimals, and rounding decimals. At the highest level, the Numeration Concepts strand encompasses a variety of objectives that could be labeled pre-algebra or simply “advanced concepts.” Included in this category are specific objectives on roots and powers, primes and composites, signed integers, and scientific notation. Because items in the Numeration Concepts strand emphasize understanding basic concepts, they are deliberately written to minimize computational burden.

Computation Processes

The Computation Processes strand includes 39 specific objectives, the second largest number among the STAR Math strands. This strand covers the four basic operations (addition, subtraction, multiplication, and division) with whole numbers, fractions, decimals, and percents. Ratios and proportions are also included in this strand. Coverage of computational skill begins with the basic facts of addition and subtraction, starting with the fact families having sums to 10, then with sums to 18. The strand progresses to addition and subtraction of two-digit and three-digit numbers without regrouping, then with regrouping. At about the same level, basic facts of multiplication and division are introduced. Then, the four operations are applied to more difficult regrouping problems with whole numbers. Fractions are first introduced by way of addition and subtraction of fractions with like denominators. These are relatively easy for students in the middle grades. However, the strand next includes operations with fractions with unlike denominators, mixed numbers, and decimal problems requiring place change, all of which are relatively difficult for students. The Computation Processes strand concludes with a series of objectives requiring operations with percents, ratios, and proportions.

Although the Computation Processes strand can be subdivided into nearly an infinite number of objectives, the STAR Math item bank provides a representative sampling of computational problems that cover the major types of problems students are likely to encounter. Indeed, the item bank does not purport to cover every conceivable computational nuance. In addition, among the more difficult problems involving computation with whole numbers, there are number combinations for which one would ordinarily use a calculator. However, it is expected that students will know how to perform these operations by hand, and hence, a number of such items are included in the STAR Math item bank.

The Numeration Concepts and Computation Processes strands are considered by many to be the heart of the basic mathematics curriculum. Students must know the four operations with whole numbers, fractions, decimals, and percents. Students must know numeration concepts to have an understanding of how the operations work, particularly for regrouping, changing denominators in fractions, and changing places with decimals and percents. As noted above, these two strands constitute the first two thirds of the STAR Math test. Mathematical development within these two strands also serves as the principal basis for instructional recommendations provided in the STAR Math Diagnostic Report.

The remaining strands comprise the latter third of the STAR Math test. This part might be labeled "applications" since many – although not all – of the objectives in this part can be considered practical applications of mathematical content and procedures. It is important to note that research conducted at the item calibration stage of STAR Math development demonstrated that the items in the various strands were strongly unidimensional, thus justifying the use of a single score for purposes of reporting.

Estimation

The Estimation strand is also designed to parallel the Computation Processes strand in terms of the types of operations required. Again, many, but not all computation objectives are reflected in this strand. Obviously, in the Estimation strand, students are not required to compute a final answer. With number combinations similar to those represented in the Computation Processes strand, students are asked to estimate an answer. To discourage students from actually computing answers, response options are generally given in round numbers. The range of numerical values used in the options is generally set so that a reasonable estimate is adequate.

Geometry

The Geometry strand in STAR Math begins with simple recognition of plane shapes and their properties. The majority of objectives in the Geometry strand concentrate on the treatment of perimeters and areas, usually covered in the middle grades, and recognition and use of parallels, intersections, and perpendiculars, covered in the middle and junior-high grades. At the more difficult levels, this strand includes application of principles about triangles and the Pythagorean theorem. Other than these latter topics, this strand does not cover the content of the typical college preparatory course in geometry.

Measurement

Although many curricular sources combine geometry and measurement in a single strand, the STAR Math test represents them separately. At the lowest level, the Measurement strand includes objectives on money, temperature, and time (clocks, days of the week, and months of the year). The strand provides coverage of both metric and customary (English) units. Metric objectives include use of the metric prefixes (milli-, centi-, etc.) and the conversion of metric and customary units. The Measurement strand also includes an objective on measurement of angles, one of the best examples of the overlap between the geometry and measurement areas.

Data Analysis and Statistics

This strand begins with simple, straightforward extraction of information from tables, bar charts, and circle graphs. In these early objectives, information needed to answer the question is given directly in the table, chart, or graph. At the next higher level of complexity, students must combine or compare two or more pieces of information in the table, chart, or graph in order to answer the question. This strand also includes several objectives related to probability and statistics. Curricular placement of probability and statistics objectives varies considerably from one source to another. In contrast, using tables, charts, and graphs is commonly encountered across a wide range of grades in nearly all mathematics curricular materials.

Word Problems

The Word Problems strand includes simple, situational applications of computations. In fact, the Word Problems strand is deliberately structured to parallel the Computation Processes strand in terms of the types of operations required. Most computation objectives are paralleled in the Word Problems strand. For all items in the Word Problems strand, students are presented with a practical problem, and to answer the item correctly, they must determine what type of computational process to use and then correctly apply that process. The reading level of the problems is kept at a low level to ensure valid assessment of ability to solve word problems.

Algebra

The final strand in the curricular structure of the STAR Math item bank is Algebra. Although algebra is generally thought of as a college preparatory course, elements of algebra are actually introduced much earlier than the high school level in the contemporary mathematics curriculum. The use of simple number sentences and the translation of word problems into equations (at a very simple level) are introduced even in the primary grades. Such objectives are included at the lowest level of the STAR Math Algebra strand. The objectives progress rapidly in difficulty to those found in the formal algebra course. These more difficult objectives include operating with polynomials, quadratic equations, and graphs of linear and non-linear functions.

Objective Clusters

The STAR Math 3.x Diagnostic Report contains two bar charts that reflect each student's performance on the Numeration Concepts and Computation Processes strands. By viewing these two charts, teachers can graphically see how each student is progressing in these two important areas. The STAR Math 3.x Diagnostic Report highlights these two strands because they form the foundation for the mathematics curriculum, especially in grades 1 through 8. According to the National Council of Teachers of Mathematics' *Principles and Standards for School Mathematics* (NCTM, 2000), "understanding numbers and operations, developing number sense, and gaining fluency in arithmetic computation form the core of mathematics education for the elementary grades" (p. 32).

The content in the Numeration Concepts and Computation Processes strands is organized in a hierarchical structure, reflecting the fact that students' mathematical development (and math curriculum) proceeds in a step-like fashion. In other words, their understanding of harder concepts is dependent upon their understanding the more basic concepts. For example, a student must first learn how to add numbers together before she is able to multiply them.

Because of this hierarchical structure and because every objective within these two strands could not be included on the STAR Math 3.x Diagnostic Report, for data reduction purposes, common objectives were grouped together, forming "objective clusters." Based on the recommendations of a mathematics content expert, the 43 Numeration Concepts objectives and the 39 Computation Processes objectives in STAR Math 2.x/3.x were grouped into 9 Computation and 8 Numeration clusters. The objectives included in each cluster in each strand are shown in Table 2.1.

Table 2.1:
Content of Objective Clusters for the STAR Math 2.x/3.x Numeration Concepts and Computation Processes Strands

Strand	Objective Cluster	Objective ID	Objective Name
Numeration Concepts	Ones	N00	Ones: Locate numbers on a number line
		NA1	Ones: Placing numerals in order
		NA2	Ones: Using numerals to indicate quantity
		NA3	Ones: Relate numerals and number words
		NA4	Ones: Use ordinal numbers

Table 2.1: (Continued)
Content of Objective Clusters for the STAR Math 2.x/3.x Numeration Concepts and Computation Processes Strands

Strand	Objective Cluster	Objective ID	Objective Name
Numeration Concepts (continued)	Tens	N01	Tens: Place numerals (10-99) in order of value
		N02	Tens: Associate numeral with group of objects
		N03	Tens: Relate numeral and number word
		N04	Tens: Identify one more/one less across decades
		N05	Tens: Understand the concept of zero
	Hundreds	N06	Hundreds: Place numerals in order of value
		N07	Hundreds: Relate numeral and number word
		N08	Hundreds: Identify place value of digits
		N09	Hundreds: Write numerals in expanded form
	Thousands	N11	Thousands: Place numerals in order of value
		N12	Thousands: Relate numeral and number word
		N13	Thousands: Identify place value of digits
		N14	Thousands: Write numerals in expanded form
	Hundred Thousands	N16	Ten thousands, hundred thousands, millions, billions: Place numerals in order of value
		N17	Ten thousands, hundred thousands, millions, billions: Relate numeral and number word
		N18	Ten thousands, hundred thousands, millions, billions: Identify place value of digits
		N19	Ten thousands, hundred thousands, millions, billions: Write numerals in expanded form

Table 2.1: (Continued)
Content of Objective Clusters for the STAR Math 2.x/3.x Numeration Concepts and Computation Processes Strands

Strand	Objective Cluster	Objective ID	Objective Name
Numeration Concepts (continued)	Fractions & Decimals	N21	Fractions and decimals: Convert fraction to equivalent fraction
		N22	Fractions and decimals: Convert fraction to decimal
		N23	Fractions and decimals: Convert decimal to fraction
		N24	Fractions and decimals: Read word names for decimals to thousandths
		N25	Fractions and decimals: Identify place value of digits in decimals
		N26	Fractions and decimals: Identify position of decimals on number line
		N27	Fractions and decimals: Identify position of fractions on number line
		N28	Fractions and decimals: Convert improper fraction to mixed number
		N29	Fractions and decimals: Round decimals to tenths, hundredths
		N30	Fractions and decimals: Relate decimals to percents
	Advanced Concepts I	N31	Advanced concepts: Determine square roots of perfect squares
		N34	Advanced concepts: Recognize meaning of exponents (2-10)
		N39	Advanced concepts: Can determine greatest common factor
		N41	Advanced concepts: Recognizes use of negative numbers

Table 2.1: (Continued)
Content of Objective Clusters for the STAR Math 2.x/3.x Numeration Concepts and Computation Processes Strands

Strand	Objective Cluster	Objective ID	Objective Name
Numeration Concepts (continued)	Advanced Concepts II	N32	Advanced concepts: Give approximate square roots of a number
		N33	Advanced concepts: Recognize the meaning of n^{th} root
		N35	Advanced concepts: Recognize meaning of negative exponents
		N36	Advanced concepts: Recognize meaning of fractional exponents
		N37	Advanced concepts: Can use scientific notation
		N38	Advanced concepts: Knows meaning of primes and composites
		N40	Advanced concepts: Can determine least common multiple
Computation Processes	Addition & Subtraction Basic Facts to 10	C01	Addition of basic facts to 10
		C02	Subtraction of basic facts to 10
	Addition & Subtraction Basic Facts to 18, No Regrouping	C03	Addition of basic facts to 18
		C04	Subtraction of basic facts to 18
		C05	Addition of three single-digit addends
		C06	Add beyond basic facts, no regrouping (2d+1d)
		C07	Subtract beyond basic facts, no regrouping (2d-1d)
	Addition & Subtraction with Regrouping	C08	Add beyond basic facts with regrouping (2d+1d, 2d+2d)
		C09	Subtract beyond basic facts with regrouping (2d-1d, 2d-2d)
		C10	Add beyond basic facts with double regrouping (3d+2d, 3d+3d)
		C11	Subtract beyond basic facts with double regrouping (3d-2d, 3d-3d)

Table 2.1: (Continued)
Content of Objective Clusters for the STAR Math 2.x/3.x Numeration Concepts and Computation Processes Strands

Strand	Objective Cluster	Objective ID	Objective Name
Computation Processes (continued)	Multiplication & Division: Basic Facts	C12	Multiplication basic facts
		C13	Division basic facts
		C14	Multiplication beyond basic facts, no regrouping (2dx1d)
	Advanced Computation with Whole Numbers	C15	Division beyond basic facts, no remainders (2d/1d)
		C16	Multiplication with regrouping (2dx1d, 2dx2d)
		C17	Division with remainders (2d/1d, 3d/1d)
		C18	Add whole numbers: any difficulty
		C19	Subtract whole numbers: any difficulty
		C21	Divide whole numbers: any difficulty
	Fractions & Decimals I	C22	Add fractions: like single digit denominators
		C23	Subtract fractions: like single digit denominators
		C33	Add decimals, place change (2+.45)
		C35	Subtract decimals, place change (5-.4)
	Fractions & Decimals II	C24	Add fractions: unlike single digit denominators
		C25	Subtract fractions: unlike single digit denominators
		C26	Multiply fractions: single digit denominators
		C27	Divide fractions: single digit denominators
		C28	Add mixed numbers
		C29	Subtract mixed numbers
		C36	Multiply decimals
		C37	Divide decimals

Table 2.1: (Continued)
Content of Objective Clusters for the STAR Math 2.x/3.x Numeration Concepts and Computation Processes Strands

Strand	Objective Cluster	Objective ID	Objective Name
Computation Processes (continued)	Percents, Ratios, & Proportions	C38	Percent A (10 is what % of 40)
		C39	Percent B (20% of 50 is what)
		C40	Percent C (30 is 50% of what)
		C41	Proportions
		C42	Ratios
	Multiplication & Division of Mixed Numbers	C30	Multiply mixed numbers
		C31	Divide mixed numbers

On the STAR Math 3.x Diagnostic Report, the shaded region of each bar chart reflects the amount of material within each strand that the student has most likely mastered. These estimates are based on the STAR Math 2.0 norming data, and mastery is defined as 70% proficient. Therefore, if a student's ability estimate suggests that she could answer 70% or more correct on a specific objective cluster, such as Hundreds, she will have "mastered" that objective cluster and that box will be shaded on her Diagnostic Report. Because the content in the strands included in the objective clusters is hierarchical, students most likely master the objective clusters in sequential order. The solid black line on the bar chart points to the objective cluster that the student is currently developing or the lowest objective that she has not mastered.

Comparison of Content in STAR Math 1.x and STAR Math 2.x/3.x

The content specifications for STAR Math 2.x/3.x were very similar to those used for STAR Math 1.x. Although the majority of the objectives are the same in the two versions, there are two main differences. First, a considerable number of easier objectives were created for STAR Math 2.x/3.x since the lowest grade targeted for STAR Math 1.x was grade 3, and STAR Math 2.x/3.x extends down to grade 1. To accommodate this downward extension, objectives typically covered in the primary grade math curriculum were added. Examples of these objectives include those on units in the Numeration Concepts strand, addition and subtraction for sums to 10 in the Computation Processes strand, and objectives on money and time in the Measurement strand. The second difference involved adding "top," or more difficult items for high-performing high school students to STAR Math 2.x/3.x. This was accomplished primarily by adding objectives to the upper level of the Algebra strand.

Rules for Writing Items

When preparing specific items to test student knowledge of the content selected for STAR Math, several item-writing rules were employed. These rules helped to shape the final appearance of the content and hence, became part of the content specifications:

- The first and perhaps most important rule was to have the item content, wording, and format reflect the typical appearance of the content in curricular materials. In some testing applications, one might want the item to look different from how the content typically appears in curricular materials. However, the goal for the STAR Math test was to have the items reflect how the content appears in curricular materials that students are likely to have used.
- Second, every effort was made to keep item content simple and to keep the required reading levels low. Although there may be some situations in which one would want to make test items appear complex or use higher levels of reading difficulty, for the STAR Math test, the intent was to simplify when possible.
- Third, efforts were made both in the item-writing and in the item-editing phases to minimize cultural loading, gender stereotyping, and ethnic bias in the items.
- Fourth, the items had to be written in such a way as to be presented in the computer-adaptive format. More specifically, items had to be presentable on the types of computer screens commonly found in schools. This rule had one major implication that influenced item presentation: artwork was limited to fairly simple line drawings, and colors were kept to a minimum.
- Finally, items were all to be presented in a multiple-choice format. Answer choices were to be laid out in either a 4x1 matrix, a 2x2 matrix, or a 1x4 matrix.

In all cases, the distracters chosen were representative of the most common errors for the particular question stem. A "not given" response option was included only for the Computation Processes strand. This option was included to minimize estimation as a response strategy and to encourage the student to actually work the problem to completion.

Computer-Adaptive Test Design

An additional level of content specification is determined by the student's performance during testing. In conventional paper-and-pencil standardized tests, items retained from the item tryout or item calibration program are organized by level. Then, each student takes all items within a given test level. Thus, the student is only tested on those mathematical operations and concepts deemed to be appropriate for his or her grade level. On the other hand, in computer-adaptive tests, such as

STAR Math, the items taken by a student are dynamically selected in light of that student's performance during the testing session. Thus, a low-performing student's knowledge of math operations may branch to easier operations to better estimate math achievement level, and high-performing students may branch to more challenging operations or concepts to better determine the breadth of their math knowledge and their math achievement level.

During an adaptive test, a student may be "routed" to items at the lowest level of difficulty within the overall pool of items, dependent upon the student's unfolding performance during the testing session. In general, when an item is responded to correctly, the student is routed to a more difficult item. When an item is answered incorrectly, the student is instead routed to an easier item. The adaptive branching procedure aims to select items such that a student is expected to have a 75 percent chance of answering each item correctly, given the student's estimated ability and the item's known difficulty. STAR Math 2.x/3.x item difficulties were determined by results of the national item calibration study, described in Chapter 3 (page 27).

A STAR Math 3.x test consists of a fixed-length, 24-item adaptive test. Students who have not taken a STAR Math 2.0 test within 180 days initially receive an item whose difficulty level is relatively easy for students at that grade level. This minimizes any effects of initial anxiety that students may have when starting the test and serves to better facilitate the student's initial reactions to the test. The starting points vary by grade level and are based on research conducted as part of the norming process described in Chapter 4 (page 36).

When a student has taken a STAR Math 2.x or 3.x test within the previous six months, the appropriate starting point is based on his or her previous test score information. Following the administration of the initial item, and after the student has entered an answer, the program determines an updated estimate of the student's math achievement level. Then, it selects the next item randomly from among all of the available items having a difficulty level that closely match this estimated achievement level. Randomization of items with difficulty values near the student's math achievement level allows the program to avoid overexposure of test items.

The items in the first part of the test (items 1 through 16) are dynamically selected from an item bank consisting of all the retained items from the Numeration Concepts and Computation Processes strands. Although the second part of the test selects items from a pool that consists of the remaining six content strands, content balancing rules ensure that every strand appropriate to the student's grade level is represented. Table 2.2 on the next page shows the content balancing design of STAR Math 2.x/3.x strands by grade.

Table 2.2:
Content Balancing Design of STAR Math 2.x/3.x Strands by Grade

	Minimum Distribution of Items by Strands											
	First 16 items (1-16)											
	Grade											
Strand	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Computation Processes	8	8	8	8	8	8	8	8	8	8	8	8
Numeration Concepts	8	8	8	8	8	8	8	8	8	8	8	8
Total	**16**	**16**	**16**	**16**	**16**	**16**	**16**	**16**	**16**	**16**	**16**	**16**
	Last 8 Items (17-24)											
	Grade											
Strand	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Algebra	0	0	0	0	0	0	0	0	2	2	2	2
Data Analysis and Statistics	1	1	1	1	1	1	1	1	1	1	1	1
Estimation	*	*	1	1	1	1	1	1	0	0	0	0
Geometry	2	2	1	1	1	2	2	2	2	2	2	2
Measurement	2	2	2	2	2	1	1	1	1	1	1	1
Word Problems	2	2	2	2	2	2	2	2	1	1	1	1
Total	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**

* Items from that strand will not be administered to students in that grade.

As can be seen in Table 2.2, all students in all grades receive eight items from Computation Processes and eight items from Numeration Concepts during the first sixteen items of the test. The specific type of question administered within these strands will vary with the student's grade level and estimated ability level. The next seven items are selected according to the student's grade level, according to Table 2.2. A zero means that no minimum criterion exists, but students may receive items from that strand if it would be consistent with the software's estimated ability level. An asterisk (*) means that items from that strand will not be administered to students in that grade.

Hence, students in grades 1 and 2 will not receive items from the Estimation strand. The final and 24th item of a STAR Math 2.x/3.x test will be selected from any available strands in Other Applications that are consistent with the student's estimated ability level.

Items that have been administered to the same student within the past six months are not available for administration. In addition, to avoid frustration, items that are intended to measure advanced mathematical concepts and operations that are more than three grade levels beyond the student's grade level, as determined by where such concepts or operations are typically introduced in math textbooks, are also not available for administration. Because the item pools make a large number of items available for selection, these minor constraints have a negligible impact on the quality of each STAR Math 2.x/3.x computer-adaptive test.

STAR Math 2.x/3.x Scoring

Following the administration of each STAR Math 2.x/3.x item, and after the student has selected a response, an updated estimate of the student's underlying math achievement level is computed based on the student's responses to all of the items administered up to that point. A proprietary Bayesian-modal item response theory estimation method is used for scoring until the student has answered at least one item correctly and at least one item incorrectly. Once the student has met this 1-correct/1-incorrect criterion, STAR Math 2.x/3.x software uses a proprietary Maximum-Likelihood IRT estimation procedure to avoid any potential bias in the Scaled Scores.

This approach to scoring enables STAR Math 2.x/3.x software to provide Scaled Scores that are statistically consistent and efficient. Accompanying each Scaled Score is an associated measure of the degree of uncertainty, called the standard error of measurement (SEM). Unlike conventional paper-and-pencil tests, the SEM values for STAR Math scores will be unique for each student dependent upon the particular items in the student's individual test and the student's performance on those items. Because the STAR Math 2.x/3.x test is computer-adaptive, however, the SEM values are relatively consistent by the end of the 24-item test.

Scaled Scores are expressed on a common scale that spans all grade levels covered by the STAR Math 2.x/3.x test. Because STAR Math 2.x/3.x software expresses Scaled Scores on a common scale, Scaled Scores are directly comparable with each other, regardless of grade level. Other scores, such as Percentile Ranks and Grade Equivalents, are derived from the Scaled Scores obtained during the STAR Math norming study described in Chapter 4 (page 36).

Chapter 3: Calibration Study and Item Analysis

In the development of STAR Math 1.0, approximately 2,450 items were prepared according to the defined STAR Math content specifications. These items were subjected to empirical tryout in 1997 in a national sample of students in grades 3 through 12. Following both traditional and item response theory (IRT) analyses of the resulting item response data, 1,434 of the items were chosen for use in the STAR Math 1.x item bank.

STAR Math 3.x uses the same item bank that was developed for STAR Math 2.0. In the development of STAR Math 2.0, about 1,100 new items were written. The new items extended the content of the STAR Math item bank to include grades 1 through 12 and expanded the algebra coverage by adding a number of new algebra objectives. Where needed, items measuring other objectives were written to supplement existing items.

All of the new items had to be calibrated on the same difficulty scale as the original STAR Math item bank. Because a number of changes in item display features were introduced with STAR Math 2.0, Renaissance Learning decided to recalibrate the STAR Math 1.x adaptive item bank simultaneously with the new items written specifically for STAR Math 2.x. During the STAR Math 2.0 Calibration Study, 2,471 items, including both the existing and the new items, were administered to a national sample of more than 44,000 students in grades 1 through 12 in the spring of 2001.

Calibration Sample

To obtain a sample that was representative of the diversity of mathematics achievement in the U.S. school population, school districts, specific schools, and individual students were selected to participate in the calibration study. The sampling frame consisted of all U.S. schools, stratified on three key variables: geographic region of the country, school size, and socioeconomic status. The STAR Math 2.0 calibration sample included students from 261 schools from 45 of the 50 United States. Tables 3.1 and 3.2 shown on the following pages present the characteristics of the calibration sample.

Table 3.1:
Sample Characteristics
STAR Math 2.0 Calibration Study – Spring 2001 (N=44,939 students)

		Students	
		National %	**Sample %**
Geographic Region	Northeast	20.4%	7.8%
	Midwest	23.5%	22.1%
	Southeast	24.3%	37.3%
	West	31.8%	32.9%
District Socioeconomic Status	Low	28.4%	30.2%
	Average	29.6%	38.9%
	High	31.8%	23.1%
	Non-public	10.2%	8.1%
School Type and District Enrollment	Public		
	<200	15.8%	24.2%
	200-499	19.1%	26.2%
	500-1999	30.2%	26.4%
	2000 or more	24.7%	15.1%
	Non-public	10.2%	8.1%

Table 3.2:
Ethnic Group and Gender Participation
STAR Math 2.0 Calibration Study – Spring 2001 (N=44,939 students)

		Students	
		National %	**Sample %**
Ethnic Group	Asian	3.9%	2.8%
	Black	16.8%	14.9%
	Hispanic	14.7%	10.3%
	Native American	1.1%	1.6%
	White	63.5%	70.4%
	Response rate	86.2%	35.7%
Gender	Female	Not available	49.8%
	Male	Not available	50.2%
	Response rate	0.0%	55.9%

In STAR Math 1.0, all test items were stored in bitmap format, and displayed on top of a bitmap image replicating a sheet of yellow graph paper. However, for STAR Math 2.x/3.x, all items were converted from bitmap format to a vector-based format. Additionally, in STAR Math 2.x/3.x, many of the new primary level math items contain bright and colorful graphics that would not reproduce well on top of the color yellow. Therefore, the yellow graph paper element common to all STAR Math 1.x items was replaced by a neutral, off-white field in STAR Math 2.0. This item field was also increased in size so graphic elements could be enlarged. Because these changes in the display format and display size could affect items' psychometric properties in STAR Math 2.x/3.x, calibration response data were collected by means of computer-administered testing, and STAR Math 1.0 items were re-calibrated along with the items newly developed for STAR Math 2.0.

Data Collection

The calibration data were collected by administering test items on-screen, with display characteristics identical to those to be implemented in the STAR Math 2.0 product. However, the calibration items were administered in forms consisting of fixed sequences of items, as opposed to the adaptive testing format.

Seven levels of test forms were constructed corresponding to varying grade levels. Because growth in mathematics is much more rapid in the lower grades, there was only one grade per level for the first four levels. As grade level increases, there is more variation among both students and school curricula, so a single test level can cover more than one grade level. Grades were assigned to test levels after extensive consultation with mathematics instruction experts, and assignments were consistent both with the STAR Math item development framework and with assignments used in other math achievement tests. To create the levels of test forms, therefore, items were assigned to grade levels such that resulting test forms sampled an appropriate range of objectives from each of the strands that are typically represented at or near the targeted grade levels. Table 3.3 describes the various test form designations used for the STAR Math 2.0 calibration study.

Table 3.3:
Test Form Levels, Grades, Numbers of Items per Form and Numbers of Test Forms STAR Math 2.0 Calibration Study – Spring 2001

Level	Grades	Items per Form	Forms	Items
A	1	36	14	152
B	2	36	22	215
C	3	36	32	310
D	4	36	34	290
E	5-6	46	36	528
F	7-8-9	46	32	516
G	10-11-12	46	32	464

Students in grades 1 through 4 (Levels A, B, C, and D) took 36-item tests consisting of three practice items and 33 actual test items. Expected testing time for these students was 30 minutes. Students in grades 5 through 12 (Levels E, F, and G) took 46-item tests consisting of three practice items and 43 actual test items. Expected testing time for these students was 40 minutes.

Items within each level were distributed among a number of test forms. Consistent with STAR Math 1.0, the content of each form was balanced between two broad categories of items: items measuring Numeration Concepts and Computation Processes and items measuring Other Applications. Each form was organized into three sections: A, B, and C. Sections A and C each consisted of approximately 40% of the test length, and contained items from both of the categories. Section A began with items measuring Numeration Concepts and Computation Processes, followed by items measuring Other Applications. Section C reversed this order, with Other Applications items preceding Numeration Concepts and Computation Processes items.

Section B comprised approximately 20% of the test length, and contained two types of anchor items. “Horizontal anchors” were common to a number of test forms at the same level, and “vertical anchors” were common to forms at adjacent levels. The anchor items were used to facilitate later analyses that placed all item difficulty parameters on a common scale.

With the exception of Levels A and G, approximately half of the vertical anchor items in each form came from the next lower level, and the other half came from the next higher level. Items chosen as vertical anchor items were selected partially based on their difficulty; items expected to be answered correctly by more than 80 percent or fewer than 50 percent of out-of-level students were not used as vertical anchor items.

Two versions of each form were used: version A and version B. Each version A form consisted of Sections A, B, and C in that order. Each version B form contained the same items, arranged in reverse order, with Section C followed by Sections B and A. The alternate forms counterbalanced the order of item presentation, as a defense against possible order effects influencing the psychometric properties of the items.

In all three test sections, items were chosen so that content was balanced at each level, with the numbers of items measuring each of the content domains roughly proportional to the distribution of items among the domains at each level.

In levels A through G combined, there were 101 unique sets of test items. Each was arranged in two alternate forms, versions A and B, that differed only in terms of item presentation order. Therefore, there was a total of 202 test forms.

Item Analysis

Following extensive quality control checks, the STAR Math 2.0 calibration item response data were analyzed by level, using both traditional item analysis techniques and item response theory (IRT) methods. For each test item, the following information was derived using traditional psychometric item analysis techniques:

- The number of students who attempted to answer the item.
- The number of students who did not attempt to answer the item.
- The percentage of students who answered the item correctly (a traditional measure of difficulty).
- The percentage of students answering each option and the alternatives.
- The correlation between answering the item correctly and the total score (a traditional measure of discrimination).
- The correlation between the endorsement of each alternative answer and the total score.

Item Difficulty

The difficulty of an item in traditional item analysis, is the percentage (or proportion) of students who answer the item correctly. This is typically referred to as the "p-value" of the item. Low p-values (such as 15%) indicate that the item is difficult since only a small percentage of students answered it correctly. High p-values indicate that the majority of students answered the item correctly and thus, the item is easy. It should be noted that the p-value only has meaning for a particular item relative to the characteristics of the sample of students who responded to it.

Item Discrimination

The traditional measure of the discrimination of an item is the correlation between the "score" on the item (correct or incorrect) and the total test score. Items that correlate highly with total test score will also tend to correlate with one another more highly and produce a test with more internal consistency. For the correct answer, the higher the correlation between the item score and the total score, the better the item is at discriminating between low-scoring and high-scoring individuals. When the correlation between the correct answer and the total test is low (or negative), the item is most likely not performing as intended. The correlation between endorsing incorrect answers and the total score should generally be low, since there should not be a positive relationship between selecting an incorrect answer and scoring higher on the overall test.

Item Response Function

In addition to traditional item analyses, the STAR Math 2.0 calibration data were analyzed using item response theory (IRT) methods. IRT methods attempt to quantitatively model what happens when a student with a specific level of ability attempts to answer a specific question. Although IRT methods encompass a family of mathematical models, the one-parameter (or Rasch) IRT model was selected for the STAR Math 2.0 data both for its simplicity and its ability to accurately model the performance of the STAR Math 2.x/3.x items.

Within IRT, the probability of answering an item correctly is a function of the student's ability and the difficulty of the item. Since IRT places the item difficulty and student ability on the same scale, this relationship can be represented graphically in the form of an item response function (IRF). Upon plotting the IRF (as represented by the solid line in Figure 3.1 on the next page), the result is an S-shaped (ogive) function. The difficulty of an item constitutes the horizontal axis; the vertical axis is the probability of a correct response.

For any specific item, the probability of answering the item correctly for students whose ability level is much less than the item's difficulty level is low. As the student's ability level increases, relative to the item's difficulty level, the probability of answering that item correctly increases until the probability nears 1.0. The midpoint, or point-of-inflection, of the IRF is the difficulty level of the item and is the point where a student with exactly the same ability level as the item's difficulty level would be expected to have a 50% chance of answering the item correctly. According to information theory, it is at or near this level that measurement of student achievement is optimal.

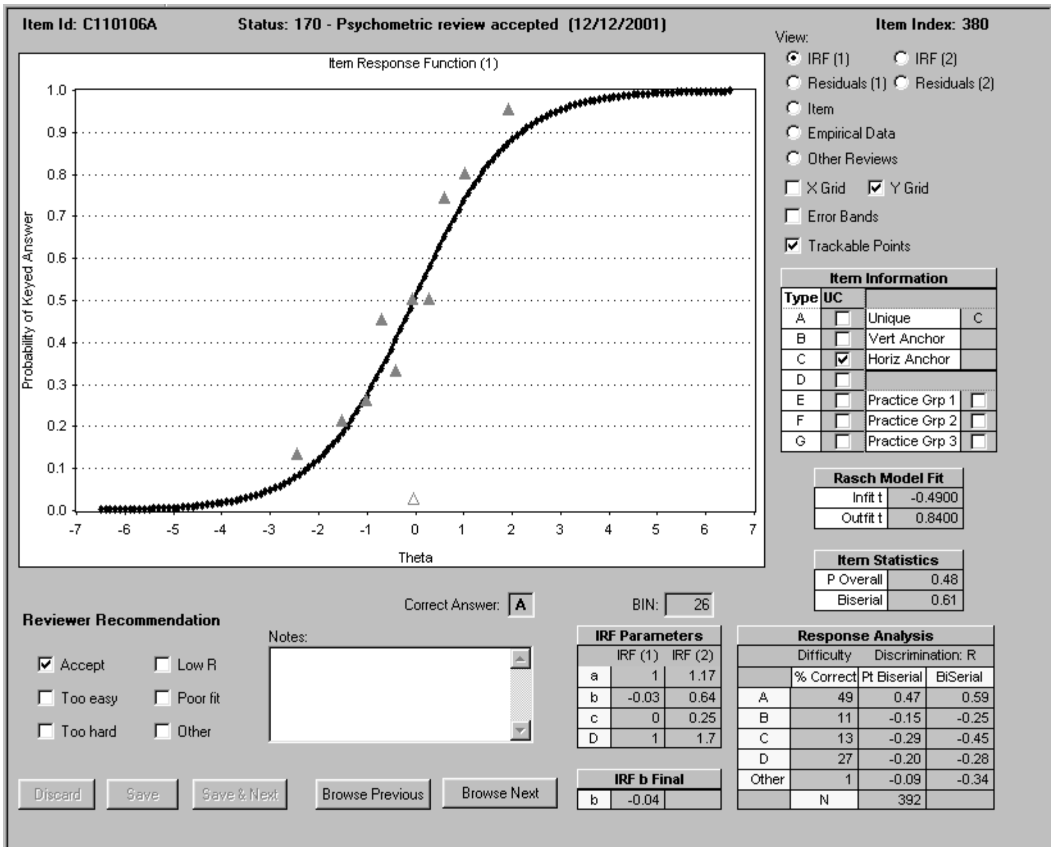


Figure 3.1: Example of Item Response Function

Calibration of test items by IRT methods estimates the IRT difficulty parameter for each test item and places all of the item parameters onto a common scale. The difficulty parameter for each item is estimated, along with measures to indicate how well the item conforms to (or "fits") the theoretical expectations of the presumed IRT model.

For purposes of the STAR Math 2.0 Calibration Study, two different "fit" measures (both unweighted and weighted) were computed, and the empirical item response function (EIRF) for the item data was determined. The EIRF is obtained by grouping students who received that item into groups with similar ability levels and then plotting the proportion of students in each group who answered the item correctly (represented by the small triangles in Figure 3.1) over the mean ability level for that group. If the IRT model is functioning well, the EIRF points should approximate the (estimated) theoretical IRF. Thus, in addition to the traditional item analysis information, the following IRT-related information was determined for each item administered during the STAR Math 2.0 Calibration Study:

- The IRT item difficulty parameter
- The unweighted measure of fit to the IRT model
- The weighted measure of fit to the IRT model
- The theoretical and empirical IRT plots

Review of Calibrated Items

Following these analyses, each test item, along with both traditional and IRT analysis information (including IRF and EIRF plots), and information about the test level, form, and item identifier were stored in a specialized item statistics database system. A panel of internal and external content reviewers then examined each item within content strands to determine whether the item met all criteria for inclusion in the bank of items that would be used in the norming version of the STAR Math 2.0 test. The item statistics database system allowed experts easy access to all available information about an item in order to interactively designate items that, in their opinion, did not meet acceptable standards for inclusion in the STAR Math 2.x item bank.

Rules for Item Retention

Items were eliminated if any of the following occurred:

- The item-total correlation (item discrimination) was less than .30.
- At least one of an item's distracters had a positive item discrimination.
- The sample size of students attempting the item was less than 300.
- The traditional item difficulty indicated that the item was too difficult or too easy.
- The item did not appear to fit the Rasch IRT model.

After each content reviewer had designated certain items for elimination, those recommendations were combined and a second review was conducted to resolve issues where there was not uniform agreement among all reviewers.

Of the initial 2,471 items administered in the STAR Math 2.0 Calibration Study, approximately 2,000 (81%) were deemed of sufficient quality to be retained for further analyses. About 1,200 of these retained items were STAR Math 1.x items.

Traditional item-level analyses were conducted again on the reduced data set. In these analyses, the dimensionality assumption of combining the first and second parts of the test was re-evaluated to ensure that all items could be placed onto a single scale. In the final IRT calibration, all test forms and levels were equated based on the information provided by the embedded anchor items within each test form so that the resulting IRT item difficulty parameters were placed onto a single scale spanning grades 1-12.

Chapter 4: Norming

STAR Math 3.x uses the same norms as STAR Math 2.0. This chapter describes the norming of STAR Math 2.0.

Sample Characteristics

The norming of the STAR Math 2.0 computer-adaptive test occurred in Spring 2002. To obtain a sample representative of the U.S. school population, the selection of participating schools focused on stratifying the U.S. school population based on three key variables. These variables, in increasing order of importance, included the following:

1. *Geographic Region.* Using the categories established by the National Education Association, schools were grouped into the following four regions: Northeast, Midwest, Southeast, and West.
2. *Per-Grade District Enrollment.* Statistics distributed by Market Data Retrieval (MDR), Inc. in 2001 identified public and non-public schools. Public schools were categorized into the following four groups based on their per-grade district enrollment: fewer than 200 students, 200-499 students, 500-1,999 students, and 2,000 or more students. Private schools were handled as a separate group, since this information was reported differently by MDR.
3. *Socioeconomic Status.* Using the Orshansky indicator listed in MDR, the U.S. school population was grouped into the following three approximately equal categories: high, average, or low socioeconomic status. Because socioeconomic data were not available for non-public schools, they were not included in this classification.

Although other data helped describe the norming sample more fully, the three variables described above were the basis for establishing an appropriate sampling frame. The sampling frame became a 52-cell matrix (4 regional zones x 4 public-school enrollment groups x 3 socioeconomic categories, plus 4 regional cells for non-public schools). All schools in the U.S. were categorized into one of the 52 cells.

A two-stage random sampling approach was used to select participating schools. In the first stage, schools in each cell were selected to receive invitation letters to participate in the STAR Math 2.0 Norming Study. In the second stage, within each cell, schools that responded to the invitation letter were randomly selected for participation.

In February 2002, the 399 schools that agreed to participate in the STAR Math 2.0 Norming Study received a special norming version of the software. The schools participating in the Norming Study began testing students in late February, and all schools were finished testing in mid-April, with a median testing date of March 18.

The norming version of the STAR Math 2.0 software administered tests in the same adaptive manner as the final version. The software also recorded the detailed information necessary for the norming analyses. After administering STAR Math 2.0 norming tests, participating schools returned their data to Renaissance Learning by creating an export file and either saving it to floppy disks for mail return or sending the file via e-mail or Internet upload. This information was also used for creation of score reports that were sent to all participating schools.

It is important to note that the STAR Math 2.0 norm-referenced scores are empirically based on each student having taken a computer-adaptive test, not simply using norms derived from a paper-and-pencil test administration. In addition, to ensure that students were instructed in a standardized format on how to take the STAR Math 2.0 computer-adaptive test, instructions to the students were carefully scripted and included in research kits.

One subset of the norming participants also participated in an alternate form reliability study. Students randomly selected by the software to participate in this study (n=7,517) were tested a second time during the norming test window. Since the STAR Math 2.x test precludes students from receiving any of the same test items for a period of six months, the correlation between the initial and second test is an alternate forms reliability coefficient. This reliability study is discussed more fully in Chapter 6 (page 48).

Another subset of the norming sample (n=3,186) participated in a study of the equivalence of STAR Math 2.0 and STAR Math 1.2. This study provided data on the degree of relationship between the new and old versions of the STAR Math tests, and it also provided a basis for score scale adjustments, if any were needed. Students randomly selected by the software to participate in this study were administered the STAR Math 1.2 test within a few days after taking their STAR Math 2.0 norming tests. This reliability study is also discussed more fully in Chapter 6 (page 48).

The final STAR Math 2.0 norming sample included a nationally representative mix of 29,185 students from 312 schools. (Appendix B, beginning on page 115, lists every school that participated in the development of STAR Math 2.0.) These schools represented 48 states across the United States[1]. Table 4.1 on page 38 summarizes the sample according to each of the variables used to select and refine the norming group.

1. **Note:** Students from five Canadian provinces also participated in the STAR Math 2.0 Norming Study. Their scores were not included in the norms, but were included in the reliability and equivalence studies.

In addition to the main sampling variables summarized in Table 4.1, other information about the sample schools was collected. Although it was not used to select or weight the STAR Math 2.0 norming sample, additional information about the norming sample is provided in Tables 4.2 through 4.4 (shown on the following pages). In some cases, not all participating schools provided the requested information, and the response rate is noted in Table 4.4 (page 39). The classifications by schools and students are provided since school sizes vary considerably. These tables also include national figures based on 2001 data provided by MDR, Inc.

Table 4.1:
Sample Characteristics
STAR Math 2.0 Norming Study – Spring 2002 (N=29,185 students)

		Students	
		National %	**Sample %**
Geographic Region	Northeast	20.4%	15.7%
	Midwest	23.5%	23.6%
	Southeast	24.3%	28.4%
	West	31.8%	32.3%
District Socioeconomic Status	Low	28.4%	26.6%
	Average	29.6%	32.6%
	High	31.8%	32.1%
	Non-public	10.2%	8.8%
School Type and District Enrollment	Public		
	<200	15.8%	20.7%
	200-499	19.1%	23.0%
	500-1999	30.2%	31.3%
	2000 or more	24.7%	16.3%
	Non-public	10.2%	8.8%

Table 4.2:
School Locations
STAR Math 2.0 Norming Study – Spring 2002 (N=312 U.S. Schools, 29,185 Students)

	Schools		Students	
	National %	**Sample %**	**National %**	**Sample %**
Urban	27.8%	23.1%	30.9%	24.8%
Suburban	38.3%	35.6%	43.5%	36.0%
Rural	33.2%	40.7%	24.8%	39.1%
Unclassified	0.7%	0.6%	0.7%	0.4%

Table 4.3:
Non-public Schools
STAR Math 2.0 Norming Study – Spring 2002 (N=27 U.S. Schools, 2,561 Students)

	Schools		Students	
	National %	**Sample %**	**National %**	**Sample %**
Catholic	39.7%	70.4%	51.8%	65.6%
Other	60.3%	29.6%	48.2%	34.4%

Table 4.4:
Gender and Ethnic Group Participation
STAR Math 2.0 Norming Study – Spring 2002 (N=29,185 students)

		Students	
		National %	**Sample %**
Ethnic Group	Asian	3.9%	1.6%
	Black	16.8%	16.2%
	Hispanic	14.7%	12.5%
	Native American	1.1%	1.2%
	White	63.5%	68.6%
	Response rate	86.2%	26.0%

Table 4.4: (Continued)
Gender and Ethnic Group Participation
STAR Math 2.0 Norming Study – Spring 2002 (N=29,185 students)

		Students	
		National %	**Sample %**
Gender	Female	Not available	49.8%
	Male	Not available	50.2%
	Response rate	0.0%	56.0%

Data Analysis

After the participating schools tested their students, they returned their student test data by e-mail, Internet upload, or on floppy disks for analysis. Following strict quality checks, these data were summarized into unit-interval frequency distributions of the Scaled Scores for students in each grade. For schools that participated in the alternate forms reliability study, the first STAR Math 2.0 test was used for the development of the norm-referenced scores; the second test was used only in the reliability data analysis.

In order to ensure that the STAR Math 2.0 norming sample was nationally representative at each grade level and to maximize the correspondence with the U.S. school population, the norming data were statistically weighted. Weighting factors were based on the number of students in each sampling cell and the national proportion of the U.S. school population that constituted each sampling cell. Weights were also applied to adjust for the disproportionately large percentage (67%) of Renaissance Learning customer schools among the participating schools in the sample. Schools were considered "customers" if they were using either STAR Math or Accelerated Math software products. Table 4.5 on the next page presents a comparison of unweighted and weighted Scaled Scores from the Norming Study.

The weighted student Scaled Score data for each grade level were then used to determine Percentile Ranks. Grade Equivalent (GE) scores were determined by plotting the empirical median weighted Scaled Scores at each grade level and interpolating the Scaled Scores for each corresponding time period between the adjacent empirical grade placement values. Scaled Score to Percentile Rank conversion tables for the empirical norming period are presented in Table 8.2 (page 80) of Chapter 8. The Scaled Score to Grade Equivalent conversion table is presented in Table 8.1 (page 76) of Chapter 8.

Table 4.5:
Comparison of Unweighted (U) and Weighted (W) Scaled Scores
STAR Math 2.0 Norming Study – Spring 2002 (N=29,185 Students)

	Sample Size		Scaled Score Means		Scaled Score Standard Deviations		Scaled Score Medians	
Grade	**U**	**W**	**U**	**W**	**U**	**W**	**U**	**W**
1	3076	2506	387	389	93	92	385	385
2	3188	2496	499	504	86	87	500	504
3	2971	2506	589	588	87	90	599	598
4	2979	2506	647	650	91	90	656	657
5	3259	2506	707	700	95	97	715	708
6	2553	2323	752	758	101	99	755	763
7	2894	2463	777	780	110	108	782	788
8	2594	2452	800	811	118	115	800	817
9	1767	2398	813	820	109	111	813	820
10	1551	2210	828	823	115	113	831	824
11	1411	2104	846	838	109	113	845	837
12	942	1742	853	852	114	116	854	852

The norming procedures for the STAR Math 2.0 Norming Study resulted in empirical, nationally representative scores for this computer-adaptive test. Because these norm-referenced scores correspond to the time period during which the norming phase was conducted (late February through mid-April), norm-referenced scores for each month of the standard nine-month school year (September-June) were then determined through a process of interpolation. This process involved interpolating between the adjacent empirical norms for each Scaled Score point, assuming equal growth between adjacent points in time and no growth over the summer months of July and August. This allows STAR Math to provide normative information that is most relevant, regardless of the specific time period in which schools administer the test to their students.

Chapter 5: Score Definitions

Use of Grade Placement in STAR Math 3.x

It is very important that the STAR Math software uses students' correct grade placement values when determining norm-referenced scores. The values of PR and NCE (Normal Curve Equivalent) are based not only on what Scaled Score the student achieved, but also on the grade placement of the student at the time of the test. For example, a second-grader in the seventh month with a Scaled Score of 534 would have a PR of 65, while a third-grader in the seventh month with the same Scaled Score would have a PR of 24. Thus, it is crucial that the STAR Math 3.x software contains the proper grade placement, and that any testing in July or August reflects the proper understanding of how STAR Math deals with these months in determining grade placement, described below.

Indicating the Appropriate Grade Placement

The numeric representation of a student's grade placement is based on the specific month in which he or she takes a test. Although teachers indicate a student's grade level or Math Instructional Level (MIL) using whole numbers, the STAR Math software automatically adds fractional increments to that grade based on the month of the test. To determine the appropriate increment, STAR Math considers the standard school year to run from September through June and assigns increment values of .0 through .9 to these months. The increment values for July and August depend on the school year setting:

- If teachers will use the July and August test scores to evaluate the student's math performance at the beginning of the year, in the Renaissance Place program, make sure the start date for that school year is before your testing in July and August. Grades are automatically increased by one level in each successive school year, so promoting students is not necessary. In this case, the increment value for July and August is 0.00 because these months are at the beginning of the school year.
- If teachers will use the test scores to evaluate the student's math performance at the end of the school year, make sure the end date for that school falls after your testing in July and August. In this case, the increment value for July and August is 0.99 because these months are at the end of the school year that has passed.

Table 5.1 on page 45 summarizes the increment values assigned to each month.

If your school follows the standard school calendar used in STAR Math and you will not be testing in the summer, assigning the appropriate grade placements for your students is automatic. However, if you are going to test students in July or August, whether it is for a summer program or because your normal calendar extends into these months, grade placements become an extremely important issue.

To ensure the accurate determination of norm-referenced scores when testing in the summer, you must determine whether to include the summer months in the past school year or in the next school year. Student grade levels are automatically increased in the new school year. In most cases, you can use the above guidelines.

Instructions for specifying school years and promoting students can be found in the *STAR Math Software Manual*.

Compensating for Incorrect Grade Placements

Teachers cannot make retroactive corrections to a student's grade placement by editing the grade assignments in a student's record or by adjusting the increments for the summer months after students have tested. The STAR Math software cannot go back in time and correct scores resulting from erroneous grade placement information. Thus, it is extremely important for the test administrator to make sure that the proper grade placement procedures are followed. If you discover that a student has tested with an incorrect grade placement assignment (use the Growth, Snapshot, Summary, or Test Record Report to find out the grade placement), the procedures outlined on the previous page in the discussion about Table 8.2 (page 80) can be used to arrive at corrected estimates for the student's Percentile Rank and Normal Curve Equivalent scores.

Types of Test Scores

In a broad sense, STAR Math software provides two different types of test scores that measure student performance in different ways:

- *Criterion-referenced scores* describe a student's performance relative to a specific content domain, or to a standard. Such scores may be expressed either on a continuous score scale, or as a classification. An example of a criterion-referenced score on a continuous scale is a percent-correct score, which expresses what proportion of test questions the student can answer correctly in the content domain. An example of a criterion-referenced classification is a proficiency category on a standards-based assessment: The student may be said to be "proficient" or not, depending on whether his score equals, exceeds, or falls below a specific criterion (the "standard") used to define "proficiency" on the standards-based test. The Numeration and Computation mastery classification charts in the Diagnostic Report are criterion-referenced.

- *Norm-referenced scores* compare a student's test results to the results of other students who have taken the same test. In this case, scores provide a relative measure of student achievement compared to the performance of a group of students at a given time. Percentile Ranks and Grade Equivalents are the two primary norm-referenced scores provided by STAR Math software. Both of these scores are based on a comparison of a student's test results to the data collected during the 2002 national norming study.

Scaled Score (SS)

STAR Math 3.x software creates a virtually unlimited number of test forms as it dynamically interacts with the students taking the test. In order to make the results of all tests comparable, and in order to provide a basis for deriving the norm-referenced scores, all STAR Math test scores are converted to a common scale, creating Scaled Scores. The STAR Math 3.x software does this in two steps. First, maximum likelihood is used to estimate each student's location on the Rasch ability scale, based on the difficulty of the items administered, and the pattern of right and wrong answers. Second, using a linear transformation to make all scores positive integers, the Rasch ability scores are converted to STAR Math Scaled Scores. STAR Math 3.x Scaled Scores range from 1 to 1400.

STAR Math 3.x Scaled Scores are expressed on the same scale used in the previous versions, STAR Math 1.x and 2.x. STAR Math Scaled Scores provide a single scale for measuring the math achievement of students from first through twelfth grade. In addition, STAR Math 3.x norm-referenced scores are derived from the within-grade distributions of Scaled Scores in the STAR Math 2.0 norms group.

Norm-referenced scores compare a student's test results to the results of other students at the same grade placement who have taken the same test. Therefore, these scores provide a relative measure of student achievement compared to the performance of a group of students at a given time. Grade Equivalent (GE), Percentile Rank (PR) and Normal Curve Equivalent (NCE) scores are the norm-referenced scores available in STAR Math. All three of these scores are based on a comparison of a student's test results to the data collected during the 2002 STAR Math 2.0 national Norming Study and are described in the following paragraphs.

Grade Equivalent (GE)

A Grade Equivalent (GE) indicates the normal grade placement of students for whom a particular score is typical. If a student receives a GE of 10.0, this means that the student scored as well on STAR Math as did the typical student at the beginning of grade 10. It does not necessarily mean that the student has mastered math objectives at a tenth-grade level, only that he or she obtained a Scaled Score as high as the average beginning tenth-grade student in the norms group.

GEs in STAR Math 3.x range from 0.0 to 12.9+. Because the GE scale expresses individual "months" in tenths, the scale does not cover the summer months. Table 5.1 below indicates how the decimalized GE tenths correspond to the various calendar months. Since the norming of STAR Math 2.0 took place during the seventh month of the school year, GEs ending in .7 are empirically based; in other words, they provide conversions based on actual normative medians. All other portions of the scale are formed by fitting a curve to the grade-by-grade medians and finding Scaled Scores that fit the curve. Table 8.1 on page 76 contains the Scaled Score to GE conversions.

Table 5.1:
Incremental Grade Placement Values Per Month

Month	Decimal Increment
July	0.0 or 0.99 (depends on the school year entered)
August	0.0 or 0.99 (depends on the school year entered)
September	0.0
October	0.1
November	0.2
December	0.3
January	0.4
February	0.5
March	0.6
April	0.7
May	0.8
June	0.9

The GE scale is not an equal-interval scale. For example, an increase of 50 Scaled Score points might represent only three or four months of GE change at the lower grades, but this same increase in Scaled Scores may signify over a year of GE change in the high school grades. This occurs because student growth in math proficiency (and other academic areas) is not linear; proficiency develops much more rapidly in the lower grades than in the middle to upper grades. Consideration of this phenomenon should be made when averaging GE scores, especially those spanning two or more grades.

Comparing STAR Math GEs With Those From Conventional Tests

Because STAR Math adapts to the proficiency level of the student being tested, the GE scores that STAR Math provides are more consistently accurate across the achievement spectrum than those provided by conventional paper-and-pencil test instruments. In addition, Grade Equivalent scores obtained using conventional test instruments are less accurate when a student's grade placement and GE score differ markedly. It is not uncommon for a fourth-grade student to obtain a GE score of 8.9 when using a conventional test instrument. However, this does not necessarily mean that the student is performing at a level typical of an end-of-year eighth-grader. More likely, it means that the student answered all, or nearly all, of the items correctly on the conventional test and thus performed beyond the range of the fourth-grade test.

On the other hand, STAR Math GE scores are more consistently accurate, even as a student's achievement level deviates from the level of grade placement. A student may be tested on any level of material up to three grade levels above grade placement, depending upon his or her actual performance on the test. Throughout a STAR Math test, students are tested on items of an appropriate level of difficulty, based on their individual level of achievement.

Percentile Rank (PR)

Percentile Rank (PR) scores indicate the percentage of students in the same grade and at the same point of time in the school year who obtained scores lower than the score of a particular student. In other words, Percentile Ranks show how an individual student's performance compares to that of his or her same-grade peers on the national level. For example, a Percentile Rank of 85 means that the student is performing at a level that exceeds 85% of other students in that grade at the same time of the year. PRs range from 1 to 99.

The PR scale is not an equal-interval scale. For example, a grade placement of 7.7 and a STAR Math 3.x Scaled Score of 868 correspond to a PR of 80, and, using the same grade placement, a STAR Math 3.x Scaled Score of 911 corresponds to a PR of 90. Thus, a difference of 43 Scaled Score points represents a 10-point difference in PR. However, for the same grade placement of 7.7, a STAR Math 3.x Scaled Score of 788 corresponds to a PR of 50, and a STAR Math 3.x Scaled Score of 812 corresponds to a PR of 60. While there is now only a 24-point difference in Scaled Scores, there is still a 10-point difference in PR. For this reason, PR scores should not be averaged or otherwise algebraically manipulated. NCE scores, described below, are much more appropriate for these types of calculations.

Table 8.2 on page 80 contains an abridged version of the Scaled Score to Percentile Rank conversion table that is used for STAR Math 3.x. The actual table includes data for all of the monthly grade placement values from 1.0 through 12.9. Because the norming of STAR Math 2.0 occurred in the seventh month of the school year, the seventh-month values for each grade are empirically based; these are the values in Table 8.2. The remaining monthly values were estimated by interpolating between the empirical points. Most of the tabled columns in this manual are appropriate for students who are at the beginning of the school year in each grade. The table also includes a column representing students who are just about to graduate from high school.

Table 8.2 can be used to estimate PR values for tests that were taken when the grade placement value of a student was incorrect (see "Use of Grade Placement in STAR Math 3.x" on page 42 for more information). One always has the option of correcting the grade placement for the student, if the error is caught right away, and then having the student retest. However, the correction technique using this Table, illustrated on the next page in example form, is intended to provide an alternate correction procedure that does not require retesting.

To illustrate, if a grade placement error occurred because a third-grade student who tested in April was accidentally entered as a fourth-grader, his or her Percentile Rank and NCE scores will be in considerable error. In order to obtain better estimates of this student's norm-referenced scores, look in the grade 3 column in Table 8.2 and locate the student's Scaled Score or the next-higher value in the table. Next, find the PR value associated with this particular Scaled Score for a student in month 7 of third grade. Then, follow the same procedure using the grade 4 column to obtain a PR corresponding to the same Scaled Score, had the student been in month 7 of fourth grade.

Teachers can use a similar interpolation procedure to obtain PR values that correspond to scores that would have been obtained at other times throughout the school year. This procedure, however, is only an approximation technique designed to compensate for grossly incorrect scores that result from a student testing while his or her grade placement was incorrectly specified. A slightly better technique involves finding the PR values in Table 8.2, converting them to NCE values using Table 8.3 (page 84), interpolating between the NCE values, and then converting the interpolated NCE value back to a PR value using Table 8.4 (page 88).

Normal Curve Equivalent (NCE)

Normal Curve Equivalents (NCEs) are scores that have been scaled in such a way that they have a normal distribution, with a mean of 50 and a standard deviation of 21.06 in the normative sample for a specific grade, for a given test. Because NCEs range from 1 to 99, they appear similar to Percentile Ranks, but they have the advantage of being based on an equal interval scale. That is, the difference between two successive scores on the scale has the same meaning throughout the scale. Because of this feature, NCEs are useful for purposes of statistically manipulating norm-referenced test results, such as interpolating test scores, calculating averages, and computing correlation coefficients between different tests. For example, in STAR Math 3.x score reports, average Percentile Ranks are obtained by first converting the PR values to NCE values, averaging the NCE values, and then converting the average NCE back to a PR.

Table 8.3 (page 84) provides the NCEs corresponding to integer PR values and facilitates the conversion of PRs to NCEs. Table 8.4 (page 88) provides the conversions from NCE to PR. The NCE values are given as a range of scores that convert to the corresponding PR value.

Chapter 6: Reliability and Measurement Precision

Reliability is a measure of the degree to which test scores are consistent across repeated administrations of the same or similar tests to the same group or population. To the extent that a test is reliable, its scores are free from errors of measurement. In educational assessment, however, some degree of measurement error is inevitable. One reason for this is that a student's performance may vary from one occasion to another. Another reason is that variation in the content of the test from one occasion to another may cause scores to vary.

In a computer-adaptive test such as STAR Math 3.x, content varies from one administration to another, and it also varies according to the level of each student's performance. Another feature of computer-adaptive tests based on Item Response Theory (IRT) is that the degree of measurement error can be expressed for each student's test individually.

The STAR Math 3.x test provides two ways to evaluate the reliability of its scores: reliability coefficients, which indicate the overall precision of a set of test scores, and conditional standard errors of measurement (SEM), which provide an index of the degree of error in an individual test score. A reliability coefficient is a summary statistic that reflects the average amount of measurement precision in a specific examinee group or in a population as a whole. In STAR Math 3.x, the SEM is an estimate of the unreliability of each individual test score. While a reliability coefficient is a single value that applies to the overall test, the magnitude of the SEM may vary substantially from one person's test score to another.

This chapter presents three different types of reliability coefficients: generic reliability, split-half reliability, and alternate forms reliability. This is followed by statistics on the conditional standard error of measurement of STAR Math 3.x test scores.

Generic Reliability

Test reliability is generally defined as the proportion of test score variance that is attributable to true variation in the trait the test measures. This can be expressed analytically as:

$$reliability = 1 - \frac{\sigma^2_{error}}{\sigma^2_{total}}$$

where σ^2_{error} is the variance of the errors of measurement, and σ^2_{total} is the variance of the test scores. In STAR Math 3.x, the variance of the test scores is easily calculated from Scaled Score data. The variance of the errors of measurement may be estimated from the conditional standard error of measurement (SEM) statistics that accompany each of the IRT-based test scores, including the Scaled Scores, as depicted on the next page.

$$\sigma^2_{error} = \frac{1}{n}\sum_{n} SEM_i^2$$

where the summation is over the squared values of the reported SEM for students i = 1 to n. In each STAR Math 3.x test, SEM is calculated along with the IRT ability estimate and Scaled Score. Squaring and summing the SEM values yields an estimate of total squared error; dividing by the number of observations yields an estimate of mean squared error, which in this case is tantamount to error variance. "Generic" reliability is then estimated by calculating the ratio of error variance to Scaled Score variance, and subtracting that ratio from 1.

Using this technique with the STAR Math 2.0 norming data resulted in the generic reliability estimates shown in the rightmost column of Table 6.1 (page 51). Because this method is not susceptible to error variance introduced by repeated testing, multiple occasions, and alternate forms, the resulting estimates of reliability are generally higher than the more conservative alternate forms reliability coefficients. These generic reliability coefficients are, therefore, plausible upper bound estimates of the actual reliability of the STAR Math 3.x computer-adaptive test.

While generic reliability does provide a plausible estimate of measurement precision, it is a theoretical estimate, as opposed to traditional reliability coefficients, which are more firmly based on item response data. Traditional internal consistency reliability coefficients such as Cronbach's alpha and Kuder-Richardson Formula 20 (KR-20) cannot be calculated for adaptive tests. However, an estimate of internal consistency reliability can be calculated using the split-half method. This is discussed in the next section.

Split-Half Reliability

In classical test theory, before the advent of digital computers automated the calculation of internal consistency reliability measures such as Cronbach's alpha, approximations such as the split-half method were sometimes used. A split-half reliability coefficient is calculated in three steps. First, the test is divided into two halves, and scores are calculated for each half. Second, the correlation between the two resulting sets of scores is calculated; this correlation is an estimate of the reliability of a half-length test. Third, the resulting reliability value is adjusted, using the Spearman-Brown formula, to estimate the reliability of the full-length test.

In internal simulation studies, the split-half method provided accurate estimates of the internal consistency reliability of adaptive tests, and so it has been used to provide estimates of STAR Math 3.x reliability. These split-half reliability coefficients are independent of the generic reliability approach discussed above and more firmly grounded in the item response data. The fifth column of Table 6.1 (page 51) contains split-half reliability estimates for STAR Math 3.x, calculated from the Norming Study data.

Alternate Form Reliability

Another method of evaluating the reliability of a test is to administer the test twice to the same examinees. Next, a reliability coefficient is obtained by calculating the correlation between the two sets of test scores. This is called a retest reliability coefficient if the same test was administered both times, and an alternate forms reliability coefficient if different, but parallel, tests were used.

This approach was used for STAR Math 2.0, as part of the Norming Study, and the results are presented in the third column of Table 6.1 (page 51). Participating schools were asked to administer two norming tests, each on a different day, to about one fourth of the overall sample. Figure 6.1 is a scatterplot of their scores. This resulted in an alternate forms reliability subsample of more than 7,000 students who took different forms of the 24-item STAR Math 2.0 norming test. The interval between the first and second tests averaged four days. The interval varied widely, however. For example, in some cases both tests were given on the same day; in other cases, the interval ranged from one to as many as 40 days.

Errors of measurement due to both content sampling and temporal changes in individuals' performance can affect alternate forms reliability coefficients, usually making them appreciably lower than internal consistency reliability coefficients. In addition, any growth in the trait that takes place in the interval between tests can also lower the correlation. The actual reliability of STAR Math 3.x is probably higher than the alternate forms estimates presented in Table 6.1 indicate.

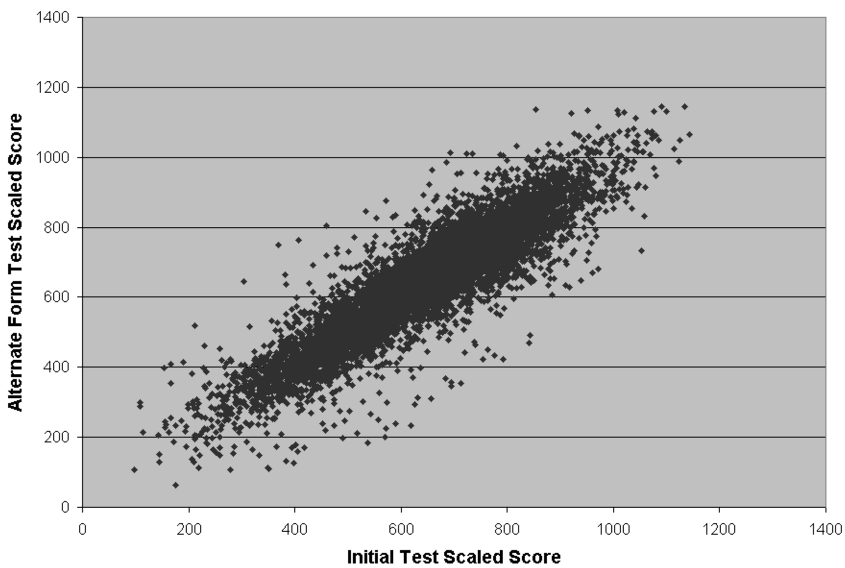


Figure 6.1: Scatterplot of Test Scores from the STAR Math 2.0 Norming Alternate Forms Reliability Study

Table 6.1 lists the detailed results of the generic, split-half, and alternate forms reliability analyses of STAR Math 2.0 Scaled Scores, both overall and by grade. The split-half and generic reliability estimates, which are based on the entire STAR Math 2.0 norms sample of 29,228[1] students, are very similar to one another, with the split-half values generally slightly lower. In the overall sample, these reliability estimates were approximately .94. By grade, they range from .78 to .88, with a median of .85.

The alternate forms reliability estimates are based on the 7,517 students who participated in the reliability study, about one fourth of the norms sample. In the overall sample, the alternate forms reliability estimates were approximately .91. By grade, the values ranged from approximately .72 to .80, with a median value of .74.

Table 6.1:
Reliability Estimates by Grade from the Norming Study – STAR Math 2.0 Scaled Scores

Grade	N	Alternate Forms Reliability	N	Split Half Reliability	Generic Reliability
1	745	0.731	3,076	0.824	0.834
2	866	0.753	3,193	0.777	0.790
3	853	0.741	2,972	0.781	0.798
4	840	0.733	2,981	0.790	0.813
5	813	0.789	3,266	0.803	0.826
6	729	0.734	2,555	0.836	0.838
7	698	0.721	2,896	0.857	0.864
8	714	0.736	2,598	0.877	0.876
9	381	0.793	1,771	0.856	0.862
10	304	0.799	1,556	0.874	0.877
11	255	0.756	1,419	0.865	0.868
12	191	0.722	945	0.882	0.872
Overall	**7,389**	**0.908**	**29,228**	**0.944**	**0.947**

1. There were 29,228 cases in the STAR Math 2.0 norms sample; 43 with outlier scores were not included in the norms calculations, but were included in the reliability calculations.

Reliability, Precision

Standard Error of Measurement

When interpreting any educational test scores, the test user must bear in mind that the scores include some degree of error. The size of the test score reliability coefficient provides an indication of the overall magnitude of that error. The standard error of measurement (SEM) arguably provides a measure that is more useful for score interpretation, as the SEM is expressed in the same units used to express the test score.

For the STAR Math 3.x Scaled Score, a conditional SEM is calculated for each individual, and the value of the SEM is included in the score reports, either explicitly or graphically.

In the following section, aggregate SEMs are presented. For the Scaled Score, these SEMs represent averages, overall and by grade. Because the conditional SEMs vary systematically by Scaled Score, the individual SEMs in the STAR Math 3.x score reports are more useful for score interpretation; the averages presented here are for purposes of test evaluation.

Scaled Score SEMs

The STAR Math 3.x software calculates the SEM for each individual. This statistic is called the "conditional SEM" as it is conditional on the value of the Scaled Score. Conditional SEMs vary from one student to another, and the interpretation of individual scores should be based on the student's own SEM value. However, for purposes of summarizing the measurement precision of STAR Math 3.x, average conditional SEM values are presented below. As the SEM estimates may vary with ability level, these SEM estimates will be tallied separately for each grade, as well as overall.

Table 6.2 on the next page contains means and standard deviations of the STAR Math 3.x Scaled Score conditional SEMs, overall and by grade, for the STAR Math 2.0 norms sample. The aggregate mean SEM value was 40, averaged over all grades. Within-grade averages range from 37 at grade 1 to 42 at grade 12.

Table 6.2:
STAR Math 2.0 Standard Error of Measurement of Scaled Scores

Grade	N	Conditional SEM	
		Mean	S.D.
1	3,076	37	5.1
2	3,193	40	4.6
3	2,972	39	3.8
4	2,981	39	3.9
5	3,266	41	4.5
6	2,555	41	4.9
7	2,896	41	5.1
8	2,598	41	5.5
9	1,771	41	5.6
10	1,556	42	6.4
11	1,419	42	6.0
12	945	42	6.6
Overall	**29,228**	**40**	**5.2**

Chapter 7: Validity

The key concept used to judge an instrument's usefulness is its validity. The validity of a test is the degree to which it assesses what it claims to measure. Determining the validity of a test is a difficult process because there are actually many aspects of validity that can be examined. For example, the content validity of the test deals with the relevance of the questions, strands, and objectives sampled by the test. These content validity issues were discussed in detail in Chapter 2 (page 13) and were an integral part of the design and construction of the STAR Math test. Construct validity, addressed in this chapter, includes the extent to which a test measures the construct that it claims to be assessing.

Establishing construct validity involves the use of data and other information external to the test instrument itself. For example, the STAR Math test claims to provide an estimate of a child's mathematical achievement level for use in placement. Therefore, demonstration of STAR Math's construct validity rests on the evidence that the test in fact provides such an estimate.

There are a number of ways to demonstrate this. One method includes examining the relationship between students' STAR Math Scaled Scores and their grade levels. Since mathematical ability varies significantly within and across grade levels and improves as a student's grade level increases, STAR Math data should demonstrate these anticipated relationships. Table 4.5 (page 41) in Chapter 4 shows a consistent pattern of grade over grade increases in average STAR Math 2.0 Scaled Scores. As STAR Math 3.x and 2.0 are psychometrically identical, this pattern is consistent with the proposition that the STAR Math 2.x/3.x test effectively measures the mathematics achievement of students.

Another source of evidence for construct validity is the relationship between students' STAR Math scores and their scores on other measures of mathematics achievement. If it is a valid assessment, the STAR Math test should correlate highly with other accepted procedures and measures that are used to determine mathematics achievement level. Among other things, students' STAR Math scores should correlate highly with their scores on other established tests of mathematics proficiency and achievement. Additionally, these scores should be highly related to teachers' assessments of their students' proficiency in mathematics.

In the remainder of this chapter, validity evidence of two kinds will be presented. First, data that demonstrate a strong and positive correlation between STAR Math 2.0 scores and scores on other standardized tests will be presented. Second, data that show a strong degree of relationship between STAR Math 2.0 scores and teacher ratings of their students' proficiency in selected math skills will be presented. All evidence supporting the validity of STAR Math 2.0 applies perforce to STAR Math 3.x.

Relationship of STAR Math 2.0 Scores to Scores on Other Tests of Mathematics Achievement

The *STAR Math 1.x Technical Manual* listed correlations between scores on that test and those on a number of other standardized measures of math achievement, obtained in 1998 for more than 9,000 students who participated in STAR Math 1.0 norming. The standardized tests included a variety of well-established instruments including the California Achievement Test (CAT), the Comprehensive Test of Basic Skills (CTBS), the Iowa Tests of Basic Skills (ITBS), the Metropolitan Achievement Test (MAT), the Stanford Achievement Test, and several statewide tests.

During the 2002 norming of STAR Math 2.0, scores on other standardized tests were obtained for more than 10,000 additional students. All of the standardized tests listed above were included, plus others such as Northwest Evaluation Association (NWEA) and TerraNova. Scores on state assessments from the following states were also included: Connecticut, Delaware, Florida, Georgia, Kentucky, Indiana, Illinois, Maryland, Michigan, Mississippi, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, Texas, Virginia, and Washington. The extent that the STAR Math 2.0 test correlates with these tests provides support for its construct validity. That is, strong and positive correlations between STAR Math 2.0 and these other instruments provide support for the claim that STAR Math 2.x effectively measures mathematics achievement.

Tables 7.1and 7.2 (page 56 and page 60) present the correlation coefficients between the scores on the STAR Math 2.0 test and each of the other test instruments for which data were received. Table 7.1 displays "concurrent validity" data, that is, correlations between STAR Math 2.0 Norming Study test scores and other tests administered at close to the same time. Tests listed in Table 7.1 were administered during the Spring of 2002, the same quarter in which the STAR Math 2.0 Norming Study took place. Table 7.2 displays all other correlations of STAR Math 2.0 norming tests and external tests; the external test scores were administered at various times prior to Spring 2002, and were obtained from student records.

Each table is presented in two parts, A and B. Part A presents validity coefficients for grades 1 through 6, and part B presents the validity coefficients for grades 7 through 12. The bottom of each table presents a grade-by-grade summary, including the total number of students for whom test data were available, the number of validity coefficients for that grade, and the average value of the validity coefficients. The within-grade average concurrent validity coefficients varied from .47 to .79, with an overall average of .65. The other validity coefficient within-grade averages varied from .56 to .70; the overall average was .63.

While these validity coefficients are high, they probably underestimate the actual correlations between the STAR Math 2.0 test and the other standardized tests of mathematics achievement. The actual relationship between the STAR Math 2.0 test and these other tests is probably higher than these estimates indicate, for several reasons. First, the standardized test scores reported were from tests administered at points in time that were different from the administration of the STAR Math 2.0 test; generally, the degree of correlation between two test scores decreases as the interval between test administrations increases. Second, most of these estimates are based on data from intact classrooms, and some restriction of the range of math achievement is to be expected with scores from intact classrooms. Range restriction is well known to attenuate correlation coefficients, as are transcription errors and other clerical errors. Third, the collection of the standardized test scores for the validity analyses involved a manual process of teachers transcribing scores for students onto forms printed by the STAR Math 2.0 norming version. Although several safeguards to reduce sources of error were implemented, this procedure was not immune to data collection errors.

The process of establishing the validity of a test is laborious, and it usually takes a significant amount of time. As a result, the validation of the STAR Math 2.x test is an ongoing activity, with the goal of establishing evidence of the test's validity for a variety of settings and students. STAR Math 1.x, 2.x, and 3.x users who collect relevant data are encouraged to contact Renaissance Learning.

Since correlation coefficients are available for many different test editions, forms, and dates of administration, many of the tests have several validity coefficients associated with them. Where test data quality could not be verified, and when sample size was very small, those data were omitted from the tabulations. Correlations were computed separately on tests according to the unique combination of test edition/form and time when testing occurred. Testing data for other standardized tests administered prior to Spring 1998 were excluded from the validity analyses.

In general, these correlation coefficients reflect very well on the validity of the STAR Math 2.x test as a tool for placement in mathematics. In fact, the correlations are similar in magnitude to the validity coefficients of these measures with each other. These validity results, combined with the supporting evidence of reliability and minimization of SEM estimates for the STAR Math 2.x test, provide quantitative demonstration of how well this innovative instrument in mathematics achievement assessment performs.

Table 7.1-A: Concurrent Validity
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered in Spring 2002, Grades 1-6, (n) = Sample Size
Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

				1		2		3		4		5		6	
Test	Version	Date	Score	n	r	n	r	n	r	n	r	n	r	n	r
California Achievement Test															
CAT	5th Ed.	S 02	NCE	–	–	–	–	17	0.50*	–	–	–	–	–	–
Comprehensive Test of Basic Skills															
CTBS	A-13	S 02	SS	–	–	–	–	–	–	–	–	21	0.66*	–	–
CTBS		S 02	NCE	–	–	–	–	–	–	–	–	–	–	32	0.65*

Table 7.1-A: Concurrent Validity (Continued)
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered in Spring 2002, Grades 1-6, (n) = Sample Size
Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	1		2		3		4		5		6	
				n	r	n	r	n	r	n	r	n	r	n	r
Iowa Tests of Basic Skills															
ITBS	Form A	S 02	NCE	–	–	–	–	–	–	50	0.66*	79	0.72*	–	–
ITBS	Form K	S 02	SS	–	–	–	–	–	–	–	–	–	–	70	0.69*
ITBS	Form L	S 02	NCE	–	–	7	0.78*	23	0.57*	17	0.70*	21	0.66*	–	–
ITBS	Form M	S 02	NCE	14	0.56*	11	0.58	–	–	–	–	–	–	–	–
ITBS	Form M	S 02	SS	–	–	–	–	17	0.72*	–	–	–	–	–	–
McGraw Hill Mississippi/Criterion Referenced															
McGraw		S 02	SS	–	–	–	–	–	–	–	–	44	0.73*	–	–
Metropolitan Achievement Test															
MAT	6th Ed.	S 02	NCE	69	0.55*	–	–	–	–	–	–	–	–	–	–
MAT	8th Ed.	S 02	SS	–	–	–	–	–	–	38	0.83*	–	–	–	–
Mississippi Curriculum Test (CTB-McGraw Hill)															
CTB	Miss	S 02	SS	–	–	–	–	–	–	10	0.62	–	–	–	–
North Carolina End of Grade															
NCEOG		S 02	NCE	–	–	–	–	70	0.60*	–	–	–	–	–	–
NCEOG		S 02	SS	–	–	–	–	62	0.73*	–	–	–	–	–	–
Oregon State Assessment															
Oregon		S 02	SS	–	–	–	–	–	–	73	0.65*	–	–	–	–
Pennsylvania System of School Assessment															
PSSA		S 02	SS	–	–	–	–	–	–	–	–	–	–	62	0.76*
Stanford Achievement Test															
SAT-9		S 02	NCE	–	–	113	0.56*	39	0.83*	46	0.54*	103	0.70*	49	0.65*
SAT-9		S 02	SS	20	0.76*	16	0.68*	18	0.59*	19	0.57*	71	0.49*	84	0.62*
TerraNova															
TerraNova		S 02	NCE	7	0.66	14	0.46	125	0.68*	18	0.67*	17	0.79*	15	0.64*

Validity

Table 7.1-A: Concurrent Validity (Continued)
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered in Spring 2002, Grades 1-6, (n) = Sample Size
Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	1		2		3		4		5		6	
				n	r	n	r	n	r	n	r	n	r	n	r
Summary															
Grade(s)	**All**			**1**		**2**		**3**		**4**		**5**		**6**	
Number of students	1581			110		161		371		271		356		312	
Number of coefficients	38			4		5		8		8		7		6	
Average validity				0.63		0.61		0.65		0.66		0.68		0.67	
Overall average	0.65														

Table 7.1-B: Concurrent Validity
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered in Spring 2002, Grades 7-12, (n) = Sample Size
Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	7		8		9		10		11		12	
				n	r	n	r	n	r	n	r	n	r	n	r
Florida Comprehensive Assessment Test															
FCAT		S 02	NCE	–	–	–	–	–	–	51	0.64*	57	0.66*	38	0.75*
Iowa Tests of Basic Skills															
ITBS	Form M	S 02	SS	37	0.40*	–	–	–	–	–	–	–	–	–	–
Michigan Comprehensive Assessment Test															
MCAS		S 02	SS	–	–	–	–	–	–	–	–	112	0.66*	–	–
New Standards Reference Mathematics Exam (Rhode Island)															
NSRME	RI	S 02	SS	–	–	–	–	–	–	–	–	67	0.67*	9	0.66

Table 7.1-B: Concurrent Validity (Continued)
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered in Spring 2002, Grades 7-12, (n) = Sample Size

Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	7		8		9		10		11		12	
				n	r	n	r	n	r	n	r	n	r	n	r
Ohio Proficiency Test															
Ohio		S 02	SS	–	–	–	–	23	0.67*	26	0.40*	24	0.77*	24	0.69*
Otis Lennon School Ability Test															
OLSAT		S 02	NCE	–	–	–	–	–	–	12	0.36	13	0.91*	6	0.72
Palmetto Achievement Challenge Test 2001															
PACT	2001	S 02	SS	–	–	161	0.72*	–	–	–	–	–	–	–	–
Stanford Achievement Test															
SAT-9		S 02	NCE	–	–	–	–	–	–	–	–	–	–	15	0.54*
SAT-9		S 02	SS	59	0.57*	9	0.85*	–	–	–	–	–	–	–	–
Texas Assessment of Academic Skills, 2001															
TAAS	2001	S 02	TLI	–	–	–	–	163	0.69*	–	–	–	–	–	–
Summary															
Grade(s)	**All**			**7**		**8**		**9**		**10**		**11**		**12**	
Number of students	906			96		170		186		89		273		92	
Number of coefficients	19			2		2		2		3		5		5	
Average validity				0.49		0.79		0.68		0.47		0.73		0.67	
Overall average	0.65														

Table 7.2-A: Other External Validity Data
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 1-6, (n) = Sample Size

Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	1		2		3		4		5		6	
				n	r	n	r	n	r	n	r	n	r	n	r
Achievement Level (RIT) Test															
RIT		F 01	SS	–	–	–	–	–	–	–	–	–	–	150	0.69*
California Achievement Test															
CAT	5th Ed.	S 01	SS	–	–	–	–	46	0.52*	–	–	–	–	–	–
Cognitive Abilities Test															
CogAT		F 00	SS	–	–	–	–	41	0.61*	–	–	–	–	–	–
CogAT		F 01	SS	–	–	45	0.73*	–	–	–	–	–	–	–	–
Comprehensive Test of Basic Skills															
CTBS	4th Ed.	S 01	GE	–	–	–	–	–	–	43	0.67*	–	–	–	–
CTBS	A-13	S 00	NCE	–	–	–	–	–	–	65	0.60*	–	–	–	–
CTBS	A-13	S 00	SS	–	–	–	–	–	–	–	–	44	0.70*	–	–
CTBS	A-13	S 01	GE	–	–	–	–	–	–	–	–	–	–	56	0.69*
CTBS	A-13	S 01	NCE	–	–	–	–	–	–	–	–	67	0.72*	–	–
CTBS	A-13	S 01	SS	–	–	–	–	–	–	42	0.61*	–	–	–	–
Connecticut Mastery Test															
Conn	2nd	F 00	SS	–	–	–	–	–	–	–	–	35	0.51*	–	–
Conn	3rd	F 01	SS	–	–	–	–	–	–	42	0.64*	–	–	27	0.52*
Des Moines Public School (Grade 2 pretest)															
DMPS		F 01	NCE	–	–	25	0.76*	–	–	–	–	–	–	–	–
Educational Development Series															
EDS	13C	S 01	GE	–	–	–	–	30	0.69*	–	–	–	–	–	–
EDS	14C	S 00	GE	–	–	–	–	–	–	32	0.44*	–	–	–	–
EDS	15C	F 01	GE	–	–	–	–	–	–	–	–	37	0.68*	–	–
Florida Comprehensive Assessment Test															
FCAT		S 01	NCE	–	–	–	–	–	–	–	–	73	0.65*	–	–

Table 7.2-A: Other External Validity Data (Continued)
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 1-6, (n) = Sample Size

Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	1		2		3		4		5		6	
				n	r	n	r	n	r	n	r	n	r	n	r
Iowa Tests of Basic Skills															
ITBS	Form A	S 01	NCE	–	–	–	–	73	0.45*	78	0.65*	–	–	–	–
ITBS	Form A	F 01	NCE	–	–	–	–	25	0.41*	25	0.35	23	0.33	86	0.81*
ITBS	Form A	F 01	SS	–	–	–	–	–	–	–	–	–	–	73	0.64*
ITBS	Form K	F 00	SS	–	–	–	–	–	–	–	–	–	–	20	0.92*
ITBS	Form K	S 01	NCE	–	–	101	0.67*	74	0.64*	31	0.25	11	0.58	31	0.62*
ITBS	Form K	F 01	NCE	–	–	–	–	10	0.78*	16	0.78*	9	0.54	18	0.63*
ITBS	Form K	F 01	SS	–	–	–	–	–	–	–	–	75	0.77*	68	0.71*
ITBS	Form L	S 01	NCE	–	–	–	–	13	0.50	46	0.81*	13	0.73*	–	–
ITBS	Form L	S 01	SS	–	–	–	–	–	–	11	0.81*	–	–	–	–
ITBS	Form L	F 01	NCE	–	–	–	–	–	–	–	–	69	0.66*	–	–
ITBS	Form M	S 99	NCE	–	–	–	–	–	–	–	–	–	–	19	0.68*
ITBS	Form M	S 00	NCE	–	–	–	–	–	–	–	–	28	0.65*	–	–
ITBS	Form M	S 01	NCE	–	–	19	0.81*	–	–	43	0.78*	–	–	–	–
ITBS	Form M	S 01	SS	–	–	–	–	47	0.39*	32	0.55*	–	–	–	–
ITBS	Form M	F 01	NCE	5	0.88*	–	–	–	–	15	0.82*	–	–	–	–
McGraw Hill Mississippi/Criterion Referenced															
McGraw		S 01	SS	–	–	–	–	–	–	–	–	121	0.52*	–	–
Metropolitan Achievement Test															
MAT	7th Ed.	F 01	NCE	–	–	–	–	–	–	–	–	–	–	15	0.84*
Michigan Education Assessment Program															
MEAP		S 01	SS	–	–	–	–	–	–	–	–	88	0.72*	–	–
Multiple Assessment Series (Primary Grades)															
Multiple		S 01	NCE	–	–	14	0.52	19	0.54*	–	–	–	–	–	–
New York State Math Assessment															
NYSMA		S 01	SS	–	–	–	–	–	–	–	–	50	0.79*	–	–

Table 7.2-A: Other External Validity Data (Continued)
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 1-6, (n) = Sample Size

Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	1		2		3		4		5		6	
				n	r	n	r	n	r	n	r	n	r	n	r
North Carolina End of Grade															
NCEOG		F 01	SS	–	–	–	–	85	0.57*	–	–	–	–	–	–
Northwest Evaluation Association Levels Test															
NWEA		S 01	NCE	–	–	–	–	–	–	–	–	83	0.81*	64	0.78*
NWEA		F 01	NCE	–	–	–	–	50	0.56*	49	0.54*	99	0.70*	–	–
Ohio Proficiency Test															
Ohio		S 01	SS	–	–	–	–	113	0.65*	–	–	–	–	–	–
Stanford Achievement Test															
SAT-9		S 99	SS	–	–	–	–	–	–	–	–	55	0.65*	–	–
SAT-9		S 00	SS	–	–	–	–	–	–	–	–	–	–	15	0.50
SAT-9		F 00	NCE	–	–	–	–	17	0.84*	20	0.83*	–	–	–	–
SAT-9		F 00	SS	–	–	–	–	–	–	–	–	–	–	46	0.58*
SAT-9		S 01	NCE	–	–	–	–	43	0.69*	–	–	50	0.38*	–	–
SAT-9		S 01	SS	64	0.52*	–	–	–	–	58	0.41*	52	0.58*	51	0.65*
SAT-9		F 01	SS	–	–	–	–	–	–	90	0.54*	32	0.67*	24	0.57*
Tennessee Comprehensive Assessment Program, 2001															
TCAP	2001	S 01	SS	–	–	–	–	–	–	–	–	48	0.56*	–	–
TerraNova															
TerraNova		S 00	NCE	–	–	–	–	–	–	–	–	–	–	43	0.60*
TerraNova		S 00	SS	–	–	–	–	–	–	–	–	11	0.61*	–	–
TerraNova		F 00	SS	–	–	–	–	–	–	–	–	108	0.62*	–	–
TerraNova		S 01	NCE	–	–	–	–	–	–	–	–	69	0.40*	85	0.62*
TerraNova		S 01	SS	–	–	–	–	–	–	104	0.50*	62	0.59*	131	0.71*
TerraNova		F 01	NCE	–	–	58	0.38*	63	0.56*	70	0.74*	85	0.61*	–	–
Test of New York State Standards															
TONYSS		S 01	SS	–	–	–	–	55	0.75*	68	0.47*	–	–	–	–

Table 7.2-A: Other External Validity Data (Continued)
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 1-6, (n) = Sample Size
Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	1		2		3		4		5		6	
				n	r	n	r	n	r	n	r	n	r	n	r
Texas Assessment of Academic Skills															
TAAS	2001	S 01	SS	–	–	–	–	–	–	78	0.52*	–	–	–	–
TAAS	2001	S 01	TLI	–	–	–	–	–	–	–	–	–	–	82	0.42*
Virginia Standards of Learning															
Virginia		S 00	SS	–	–	–	–	–	–	–	–	24	0.73*	–	–
Washington Assessment of Student Learning															
Wash		S 00	SS	–	–	–	–	–	–	–	–	–	–	90	0.54*
Wide Range Achievement Test															
WRAT III		F 01	NCE	–	–	–	–	–	–	44	0.32*	44	0.66*	–	–
Summary															
Grade(s)	**All**			**1**		**2**		**3**		**4**		**5**		**6**	
Number of students	4996			69		262		804		1102		1565		1194	
Number of coefficients	98			2		6		17		23		29		21	
Average validity				0.70		0.65		0.60		0.59		0.62		0.65	
Overall average	0.62														

Validity

Table 7.2-B: Other External Validity Data
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 7-12, (n) = Sample Size
Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	7		8		9		10		11		12	
				n	r	n	r	n	r	n	r	n	r	n	r
American College Testing Program															
ACT		F 01	NCE	–	–	–	–	–	–	–	–	–	–	26	0.87*
California Achievement Tests															
CAT	5th Ed.	F 01	NCE	–	–	–	–	64	0.73*	–	–	–	–	–	–
CAT	5th Ed.	F 01	SS	170	0.54*	–	–	–	–	–	–	–	–	–	–
Comprehensive Test of Basic Skills															
CTBS	4th Ed.	S 00	SS	67	0.67*	75	0.73*	–	–	–	–	–	–	–	–
CTBS	A-13	S 00	SS	–	–	31	0.65*	–	–	–	–	–	–	–	–
CTBS	A-13	S 01	SS	23	0.82*	–	–	–	–	48	0.63*	–	–	–	–
Delaware Student Testing Program															
DSTP		S 01	SS	–	–	–	–	94	0.27*	–	–	–	–	–	–
Differential Aptitude Tests															
DAT	Level 1	F 01	NCE	–	–	–	–	41	0.70*	–	–	–	–	–	–
Explore Tests															
Explore		F 01	NCE	–	–	64	0.54*	–	–	–	–	–	–	–	–
Georgia High School Graduation Test															
Georgia		S 01	NCE	–	–	–	–	–	–	–	–	–	–	23	0.71*
Indiana Statewide Testing for Educational Progress															
ISTEP		F01	NCE	–	–	–	–	51	0.57*	22	0.58*	–	–	–	–
Iowa Tests of Basic Skills															
ITBS	Form A	F 01	SS	66	0.71*	–	–	–	–	–	–	–	–	–	–
ITBS	Form K	S 01	NCE	73	0.80*	18	0.52*	–	–	–	–	–	–	–	–
ITBS	Form K	F 01	NCE	6	0.72	14	0.69*	–	–	–	–	–	–	–	–
ITBS	Form L	S 01	NCE	36	0.74*	32	0.53*	–	–	19	0.67*	32	0.84*	–	–

Table 7.2-B: Other External Validity Data (Continued) STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 7-12, (n) = Sample Size

Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	7		8		9		10		11		12	
				n	r	n	r	n	r	n	r	n	r	n	r
ITBS	Form M	S 99	NCE	–	–	5	0.89*	–	–	–	–	11	0.80*	–	–
ITBS	Form M	S 00	NCE	–	–	–	–	–	–	9	0.94*	–	–	–	–
ITBS	Form M	S 01	NCE	49	0.52*	48	0.51*	–	–	–	–	–	–	–	–
Kentucky Core Content Test															
KCCT		S 01	NCE	–	–	–	–	45	0.43*	–	–	–	–	–	–
Maryland High School Placement Test															
Maryland		S 01	NCE	–	–	–	–	47	0.60*	–	–	–	–	–	–
McGraw Hill Mississippi/Criterion Referenced															
McGraw		S 01	SS	–	–	–	–	73	0.56*	–	–	–	–	–	–
Metropolitan Achievement Test															
MAT	7th Ed.	F 01	NCE	5	0.80	11	0.82*	–	–	–	–	–	–	–	–
North Carolina End of Grade Tests															
NCEOG		S 01	SS	–	–	177	0.59*	–	–	–	–	–	–	–	–
Oklahoma School Testing Program Core Curriculum Tests															
Oklahoma		S 01	SS	–	–	–	–	26	0.67*	–	–	–	–	–	–
Oregon State Assessment															
Oregon		S 01	NCE	46	0.49*	45	0.53*	–	–	–	–	–	–	–	–
PLAN															
PLAN		F 99	SS	–	–	–	–	–	–	–	–	–	–	10	0.42
PLAN		F 00	SS	–	–	–	–	–	–	–	–	40	0.28	–	–
PLAN		F 01	NCE	–	–	–	–	–	–	63	0.61*	–	–	–	–

Table 7.2-B: Other External Validity Data (Continued)
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 7-12, (n) = Sample Size
Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	7		8		9		10		11		12	
				n	r	n	r	n	r	n	r	n	r	n	r
Preliminary SAT/National Merit Scholarship Qualifying Test															
PSAT/ NMSQT	NMSQT	F 00	NCE	–	–	–	–	–	–	–	–	–	–	37	0.63*
PSAT/ NMSQT	NMSQT	F 01	NCE	–	–	–	–	–	–	–	–	72	0.64*	–	–
Stanford Achievement Test															
SAT-9		S 98	NCE	11	0.84*	–	–	–	–	–	–	–	–	–	–
SAT-9		S 99	NCE	14	0.71*	–	–	–	–	–	–	–	–	–	–
SAT-9		F 00	SS	–	–	45	0.85*	–	–	–	–	–	–	–	–
SAT-9		S 01	NCE	45	0.71*	105	0.81*	11	0.69*	–	–	–	–	–	–
SAT-9		S 01	SS	54	0.76*	109	0.69*	19	0.27	77	0.59*	67	0.76*	71	0.65*
SAT-9		F 01	SS	104	0.84*	–	–	–	–	–	–	–	–	–	–
TerraNova															
TerraNova		S 99	NCE	35	0.61*	47	0.62*	–	–	–	–	–	–	–	–
TerraNova		S 00	SS	18	0.73*	–	–	–	–	–	–	–	–	–	–
TerraNova		S 01	NCE	17	0.29	17	0.52*	–	–	–	–	–	–	–	–
TerraNova		S 01	SS	–	–	99	0.74*	–	–	–	–	–	–	–	–
TerraNova		F 01	SS	–	–	38	0.74*	–	–	–	–	–	–	–	–
Test of Achievement Proficiency															
TAP		F 01	NCE	–	–	–	–	8	0.70	7	0.70	–	–	–	–
Texas Assessment of Academic Skills, 2001															
TAAS	2001	S 01	SS	66	0.44*	69	0.33*	–	–	–	–	–	–	–	–
Virginia Standards of Learning															
Virginia		S 00	SS	25	0.71*	–	–	–	–	–	–	–	–	–	–

Table 7.2-B: Other External Validity Data (Continued)
STAR Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 7-12, (n) = Sample Size
Asterisks (*) denote correlation coefficients that are statistically significant at the .05 level.

Test	Version	Date	Score	7		8		9		10		11		12	
				n	r	n	r	n	r	n	r	n	r	n	r
Summary															
Grade(s)	**All**			**7**		**8**		**9**		**10**		**11**		**12**	
Number of students	3066			930		1049		479		245		222		141	
Number of coefficients	66			20		19		11		7		5		4	
Average validity				0.67		0.65		0.56		0.67		0.66		0.60	
Overall average	0.64														

Meta-Analysis of the STAR Math Validity Data

Meta-analysis is a set of statistical procedures that combines results from different sources or studies. When applied to a set of correlation coefficients that estimate test validity, meta-analysis combines the observed correlations and sample sizes to yield estimates of overall validity, as well as standard errors and confidence intervals, both overall and within grades. To conduct a meta-analysis of the STAR Math validity data, the 276 correlations displayed in Tables 7.1 and 7.2 were combined and analyzed using a fixed effects model for meta-analysis. The results are displayed in Table 7.3 on the next page. The table lists results for the correlations within each grade, as well as results with all twelve grades' data combined. For each set of results, the table lists an estimate of the true validity, a standard error, and the lower and upper limits of a 95 percent confidence interval for the validity coefficient.

Using the pilot study data, the overall estimate of the validity of STAR Math is .64, with a standard error of .005. The true validity is estimated to lie within the range of .63 to .65, with a 95 percent confidence level. The probability of observing the 276 correlations reported in Tables 7.1 and 7.2, if the true validity were zero, is virtually zero. Because the 276 correlations were obtained with widely different tests, and among students from twelve different grades, these results provide support for the validity of STAR Math as a measure of math skills.

Table 7.3: Results of the Meta-Analysis of STAR Math 2.0 Correlations with Other Tests

	Effect Size		95% Confidence Interval	
Grade	Validity Estimate	Standard Error	Lower Limit	Upper Limit
1	0.58	0.05	0.48	0.68
2	0.61	0.03	0.55	0.67
3	0.61	0.02	0.58	0.65
4	0.59	0.02	0.55	0.62
5	0.64	0.01	0.61	0.67
6	0.66	0.01	0.64	0.67
7	0.64	0.02	0.60	0.68
8	0.65	0.02	0.62	0.69
9	0.57	0.03	0.52	0.63
10	0.60	0.04	0.53	0.67
11	0.68	0.03	0.62	0.72
12	0.68	0.03	0.61	0.75
All Grades	0.64	0.00	0.63	0.65

Relationship of STAR Math 2.0 Scores to Teacher Ratings

In order to have a common measure of each student's math skills independent of STAR Math 2.0, Renaissance Learning constructed two 12-item checklists for teachers to use during the Norming Study. On this worksheet, teachers were asked to rate each student's ability to complete a wide range of tasks related to developing math skills. The intent of this checklist was to provide teachers with a single, brief instrument they could use to rate any student.

For simplicity, two rating forms were developed: one for grades 1 through 5, and another for grades 6 through 12. This section presents the skills rating instrument itself, its psychometric properties as observed in the Norming Study, and the relationship between student skills ratings on the instrument and their Scaled Scores on STAR Math 2.0.

The Rating Instruments

To gather ratings of math skills from teachers, these instruments were intended to specify a sequence of skills that the teacher could quickly assess for each student and were ordered such that a student who could correctly perform the n^{th} skill in the list could almost certainly perform all of the preceding skills correctly as well. Such a list, even though quite short, provided a reliable method for sorting students from first through twelfth grade into an ordered set of math skill categories.

To construct the two ratings instruments, nineteen skill-related items were written, ranked from easiest to hardest, and assembled into two rating instruments. The first twelve items – the twelve easiest skills – formed the rating instrument used for grades 1 to 5. The eighth through nineteenth items – the 12 hardest skills – made up the instrument used for grades 6 through 12.

Each teacher was asked to dichotomously rate his or her students participating in the STAR Math 2.0 Norming Study on each skill using the rating form appropriate to the student's grade. To assist with this process, the Norming Study software incorporated a feature enabling it to print a ratings worksheet for each participating grade. The printed ratings worksheet consisted of a checklist of the twelve skill-related performance tasks, pre-printed with the names of the participating students. To complete the instrument, the teacher had to simply mark, for each student, any task he/she believed the student could perform. The items forming both rating forms are shown on the following two pages.

Grade 1 – 5 Math Skills Rating Worksheet

STAR Math 2.0 Norming

For Grades 1 - 5

Sorted by: Student Name ***School Name:***

Primary Contact:

In the table below, please identify which of the following tasks each of your students can probably do correctly.

1. Identify the longest pencil among 3 pencils of different lengths.
2. Add 2 to 4.
3. State how many cents a dime is worth.
4. Determine the number that shows “ones” in 162.
5. Subtract 7 from 35.
6. Determine the number that follows in the sequence 2, 6, 10, 14, ____.
7. Divide 18 by 3.
8. Write 78,318 in expanded form.
9. Read aloud the word name for 0.914.
10. Solve the problem 4/9 + 8/9.
11. Translate the statement “36 divided by a number is 12” into an equation.
12. Divide 11,540 by 577.

Renaissance Learning, Inc. and its subsidiaries maintain high standards of confidentiality with all data acquired for research and development purposes. Renaissance Learning assures you that all school and student data derived from these activities will only be used for research and development purposes that are intended to validate and/or improve design specifications for general product release into the education market. Individual teacher and student names, grades, and ages will be kept strictly confidential; access to this data will be limited to personnel with relevant research and development responsibilities.

Grade: 1

		Mark an “X” for the tasks that each student probably can do correctly and an “O” for the tasks that each student probably cannot do correctly:												
Student No.	**Student Name**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Not Rated**
1	Bartles, Amanda													
2	Bowers, Erica													
3	Driggon, Haley													
4	Edmond, Mason													
5	Edwards, Robert													
6	Halstead, Matthew													
7	Jackson, Wesley													
8	Kendricks, Marcy													
9	Lyons, Freda													
10	Renquist, Ryan													

Grade 6 – 12 Math Skills Rating Worksheet

STAR Math 2.0 Norming

For Grades 6 - 12

Sorted by: Student Name ***School Name:***

Primary Contact:

In the table below, please identify which of the following tasks each of your students can probably do correctly.

1. Write 78,318 in expanded form.
2. Read aloud the word name for 0.914.
3. Solve the problem 4/9 + 8/9.
4. Translate the statement “36 divided by a number is 12” into an equation.
5. Divide 11,540 by 577.
6. Solve a word problem requiring the calculation of proportions.
7. Solve the problem “14 is 50% of what number.”
8. Solve a word problem requiring the calculation of 80% of 112.
9. Simplify the expression (x + 1)(x + 4).
10. Solve the equation $x^2 = 16x$.
11. Calculate vertical and supplementary angles.
12. Determine 6^{-2}.

Renaissance Learning, Inc. and its subsidiaries maintain high standards of confidentiality with all data acquired for research and development purposes. Renaissance Learning assures you that all school and student data derived from these activities will only be used for research and development purposes that are intended to validate and/or improve design specifications for general product release into the education market. Individual teacher and student names, grades, and ages will be kept strictly confidential; access to this data will be limited to personnel with relevant research and development responsibilities.

Grade: 6

		Mark an “X” for the tasks that each student probably <u>can</u> do correctly and an “O” for the tasks that each student probably <u>cannot</u> do correctly:												
Student No.	**Student Name**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Not Rated**
1	Bailey, Amanda													
2	Blake, Erica													
3	Duey, Haley													
4	Eaton, Mason													
5	Erlings, Robert													
6	Gable, Matthew													
7	James, Wesley													
8	Koore, Marcy													
9	Lipton, Freda													
10	Taylor, Ryan													

Psychometric Properties of the Skills Ratings

Teachers completed skills ratings for 17,326 of the 29,185 students in the U.S. norms group. The skills rating items were calibrated on an IRT scale using the Rasch model, with item parameters from both levels placed on a common scale. This allowed the skills ratings for students at both levels to be assigned a score on the same Rasch metric.

The resulting Rasch scores ranged from -14.47 to 11.1. The lower value corresponds to students in grades 1 to 5 rated as possessing none of the math skills, and the higher value corresponds to students in grades 6 through 12 rated as possessing all of them. Table 7.4 lists data about the psychometric properties of the rating scale, overall and by grade, including the correlations between skills ratings and STAR Math 2.0 Scaled Scores. The internal consistency reliability of the rating scale was estimated as .93, using Cronbach's alpha.

Table 7.4:
Psychometric Characteristics of the Skills Rating Scale and its Relationship to Scaled Scores, by Grade (Coefficient Alpha = 0.93)

Grade	N	Skills Rating		STAR Math 2.0 Scaled Score		Correlation of Skills Ratings and Scaled Scores
		Mean	S.D.	Mean	S.D.	
1	1,916	-6.60	2.95	385	89	0.40*
2	2,043	-3.67	2.41	503	84	0.47*
3	1,817	0.04	3.06	589	87	0.52*
4	1,820	1.26	2.83	651	90	0.58*
5	2,072	2.97	2.84	713	97	0.50*
6	1,637	5.50	2.07	763	100	0.44*
7	1,465	5.57	2.18	785	109	0.50*
8	1,639	6.96	2.50	811	117	0.54*
9	1,036	6.88	2.87	798	110	0.52*
10	688	8.78	2.38	824	119	0.38*
11	737	9.81	2.30	847	123	0.39*
12	456	10.03	2.05	876	127	0.42*
Overall	**17,326**	**2.42**	**5.60**	**672**	**177**	**0.85***

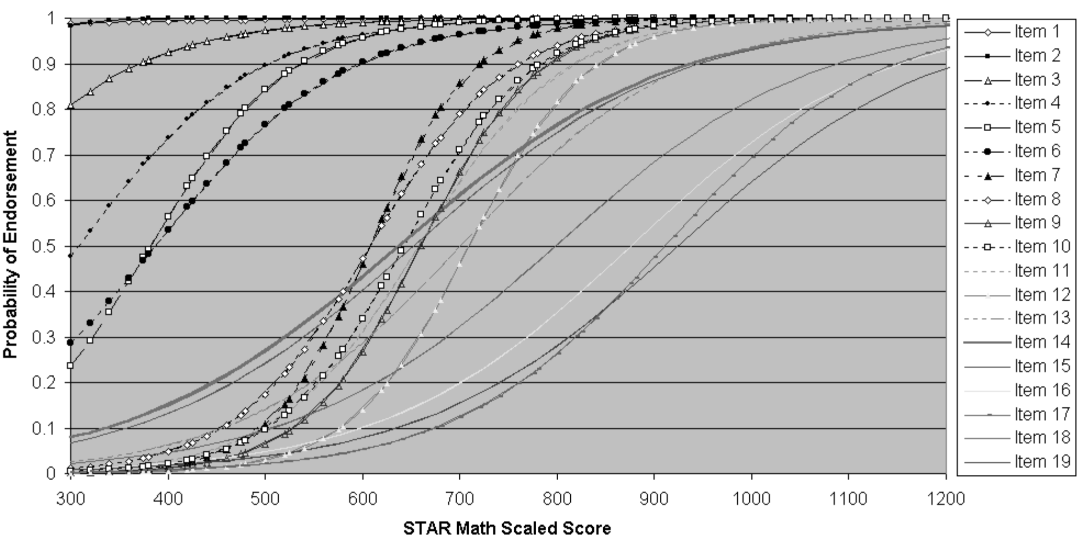
* Asterisks denote correlation coefficients that are statistically significant at the .05 level.

Relationship of STAR Math 2.0 Scaled Scores to Math Skills Ratings

As the data in Table 7.2 (page 72) show, the mean rating Scaled Scores increased directly with grade, from -6.6 at grade 1 to 10.03 at grade 12. The correlation between the skills ratings and STAR Math 2.0 Scaled Scores was significant at every grade level. The overall correlation was .85, indicating a substantial degree of relationship between the computer-adaptive STAR Math 2.x test and teachers' ratings of their students' math skills.

Figure 7.1 displays the relationships of each of the nineteen rating scale items to STAR Math 2.0 Scaled Scores. These relationships were obtained by fitting mathematical models to the response data for each of the rating items. Each of the curves in the figure is a graphical depiction of the respective model. As the curves show, the proportion of students rated as possessing each of the twelve rated skills increases with the STAR Math 2.0 Scaled Score.

Figure 7.1. The Relationship of Teachers' Ratings of Student Math Skills to STAR Math Scaled Scores



The relative positions of the curves provide one indication of the relative difficulty of the 19 rated skills. The rating items' Rasch difficulty parameters, displayed in Table 7.5 on the next page, provide a somewhat different indication; the skills rating items are listed in the table from easiest to most difficult, by Rasch difficulty. The first column of Table 7.5 indicates the relative difficulty of the nineteen rating items, where relative difficulty 1 is the easiest and 19 is most difficult. The second and third columns list the item numbers and text of the skills rating items. The fourth column lists the Rasch difficulty scale value for each item. The fifth column lists the correlations between students' ratings and their STAR Math 2.0 Scaled Scores.

Table 7.5:
The Nineteen Rating Scale Items Listed in Order of Difficulty with Rasch Difficulty Parameters

Relative Difficulty	Item	Rating Scale Item	Rasch Difficulty	Correlation with Scaled Score
1 Easiest	1	Identify the longest pencil among 3 pencils of different lengths.	-14.58	.06*
2	2	Add 2 to 4.	-14.30	.09*
3	3	State how many cents a dime is worth.	-10.28	.26*
4	4	Determine the number that shows "ones" in 162.	-7.26	.43*
5	5	Subtract 7 from 35.	-6.12	.55*
6	6	Determine the number that follows in the sequence 2, 6, 10, 14, _____.	-5.42	.49*
7	7	Divide 18 by 3.	-1.85	.71*
8	8	Write 78,318 in expanded form.	1.22	.67*
9	10	Solve the problem 4/9 + 8/9.	2.09	.70*
10	9	Read aloud the word name for 0.914.	2.51	.70*
11	11	Translate the statement "36 divided by a number is 12" into an equation.	2.59	.67*
12	12	Divide 11,540 by 577.	3.89	.68*
13	14	Solve the problem "14 is 50% of what number."	4.54	.40*
14	15	Solve a word problem requiring the calculation of 80% of 112.	4.75	.34*
15	13	Solve a word problem requiring the calculation of proportions.	5.12	.35*
16	18	Calculate vertical and supplementary angles.	6.85	.35*
17	16	Simplify the expression (x + 1)(x + 4).	8.10	.37*
18	19	Determine 6^{-2}.	9.03	.36*
19 Most Difficult	17	Solve the equation $x^2 = 16x$.	9.12	.33*

* Asterisks denote correlation coefficients that are statistically significant at the .05 level.

Notice that the first two rating scale items ("Identify the longest pencil among 3 pencils of different lengths" and "Add 2 to 4") had extremely low Rasch difficulty indices, and correlations with Scaled Scores that were near zero. As can be seen in Figure 7.1 (page 73), these items were endorsed for nearly 100% of the students, regardless of their STAR Math 2.0 Scaled Scores. As a result, they did not discriminate among students with high and low levels of developed math ability, as measured by the STAR Math 2.0 test.

Although teachers endorsed items 3 through 6 somewhat less often than items 1 and 2, they still considered these math tasks relatively easy for their students to complete. The correlations with STAR Math 2.0 Scaled Scores for items 3 through 6 were higher than those for the first two items, but still only moderate. This may have occurred because the skills associated with items 3 through 6 are almost completely mastered (defined as 80% proficiency) by a student obtaining a STAR Math 2.0 Scaled Score of 500.

Teachers' responses to items 7 through 12 suggest that their corresponding math tasks are considerably more difficult for their students to complete. This is reflected both in their Rasch difficulty parameters in Table 7.5 (page 74) and in Figure 7.1. The figure suggests that mastery of these skills occurs between 700 and 800 on the STAR Math 2.0 Score Scale. The slopes of the curves for these are all steep relative to other skills items, suggesting that these skills develop rapidly, compared to the others. The correlations between these items and Scaled Scores support this hypothesis, as items 7 through 12 show the highest correlations with STAR Math 2.0 Scaled Scores.

Items 13 through 19 measure the most difficult of the skills. This is indicated by their Rasch difficulty parameters in Table 7.5 and is also confirmed by the locations at which 80% mastery occurs, illustrated in Figure 7.1, which suggests that these skills develop much later than all others. In fact, all students may not master these skills. Moreover, all of these items have only moderate correlations with STAR Math 2.0 Scaled Scores, suggesting that growth of these skills is relatively gradual.

Chapter 8: Conversion Tables

Table 8.1:
Scaled Score to Grade Equivalent Conversions

Scaled Score	Grade Equivalent*
1 - 194	0.0
195 - 206	0.1
207 - 217	0.2
218 - 229	0.3
230 - 241	0.4
242 - 253	0.5
254 - 265	0.6
266 - 277	0.7
278 - 289	0.8
290 - 301	0.9
302 - 313	1.0
314 - 325	1.1
326 - 336	1.2
337 - 348	1.3
349 - 360	1.4
361 - 372	1.5
373 - 384	1.6
385 - 396	1.7
397 - 408	1.8
409 - 420	1.9
421 - 432	2.0
433 - 444	2.1
445 - 455	2.2
456 - 467	2.3
468 - 479	2.4
480 - 491	2.5
492 - 503	2.6

Table 8.1: (Continued)
Scaled Score to Grade Equivalent Conversions

Scaled Score	Grade Equivalent*
504 - 512	2.7
513 - 522	2.8
523 - 531	2.9
532 - 541	3.0
542 - 550	3.1
551 - 559	3.2
560 - 569	3.3
570 - 578	3.4
579 - 588	3.5
589 - 597	3.6
598 - 603	3.7
604 - 609	3.8
610 - 615	3.9
616 - 621	4.0
622 - 627	4.1
628 - 632	4.2
633 - 638	4.3
639 - 644	4.4
645 - 650	4.5
651 - 656	4.6
657 - 661	4.7
662 - 666	4.8
667 - 671	4.9
672 - 676	5.0
677 - 682	5.1
683 - 687	5.2
688 - 692	5.3
693 - 697	5.4
698 - 702	5.5
703 - 707	5.6
708 - 713	5.7

Table 8.1: (Continued)
Scaled Score to Grade Equivalent Conversions

Scaled Score	Grade Equivalent*
714 - 718	5.8
719 - 724	5.9
725 - 729	6.0
730 - 735	6.1
736 - 740	6.2
741 - 746	6.3
747 - 751	6.4
752 - 757	6.5
758 - 762	6.6
763 - 765	6.7
766 - 767	6.8
768 - 770	6.9
771 - 772	7.0
773 - 775	7.1
776 - 777	7.2
778 - 780	7.3
781 - 782	7.4
783 - 785	7.5
786 - 787	7.6
788 - 790	7.7
791 - 793	7.8
794 - 796	7.9
797 - 799	8.0
800 - 802	8.1
803 - 804	8.2
805 - 807	8.3
808 - 810	8.4
811 - 813	8.5
814 - 816	8.6
817 - 817	8.7
818 - 818	9.1

Table 8.1: (Continued)
Scaled Score to Grade Equivalent Conversions

Scaled Score	Grade Equivalent*
819 - 819	9.5
820 - 820	9.8
821 - 821	10.0
822 - 822	10.3
823 - 823	10.5
824 - 824	10.7
825 - 826	10.8
827 - 827	10.9
828 - 828	11.0
829 - 830	11.1
831 - 831	11.2
832 - 832	11.3
833 - 833	11.4
834 - 835	11.5
836 - 836	11.6
837 - 838	11.7
839 - 839	11.8
840 - 841	11.9
842 - 842	12.0
843 - 844	12.1
845 - 845	12.2
846 - 847	12.3
848 - 848	12.4
849 - 850	12.5
851 - 851	12.6
852 - 853	12.7
854 - 854	12.8
855 - 857	12.9
858 - 1400	12.9+

* **Note:** Because the range of median Scaled Scores in the STAR Math 2.0 norming study was smaller than the number of possible Grade Equivalent points, some Grade Equivalent score values in grades 8 through 10 are not listed.

Table 8.2:
Scaled Score to Percentile Rank Conversions by Grade (at month 7 in the school year)

	Grade											
PR	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	159	268	341	417	425	483	491	492	525	526	559	580
2	196	310	363	449	459	530	537	548	570	571	598	608
3	211	334	389	469	489	550	552	580	596	597	616	630
4	226	352	417	478	504	573	578	598	615	617	624	645
5	237	363	429	487	531	583	593	613	630	632	649	664
6	245	376	438	501	541	596	607	628	643	644	662	673
7	257	385	448	508	554	611	617	639	653	655	671	682
8	263	392	458	519	560	620	627	645	663	665	679	693
9	269	400	462	524	571	629	637	655	671	673	689	699
10	276	406	468	533	579	641	644	661	678	681	696	706
11	281	411	472	541	586	645	652	667	685	687	704	711
12	286	415	478	544	590	648	659	676	692	694	708	718
13	289	420	487	550	596	651	662	681	697	700	718	722
14	295	423	490	555	600	652	666	688	703	705	721	726
15	299	427	494	559	607	660	672	696	708	711	725	730
16	303	432	500	563	610	668	676	702	713	716	730	733
17	307	436	503	566	614	673	681	706	718	720	736	736
18	311	438	507	570	620	678	687	711	722	725	740	744
19	314	441	511	574	625	678	689	715	726	729	744	750
20	318	443	514	578	629	683	694	721	731	733	747	755
21	322	446	519	581	632	685	696	726	734	737	748	759
22	324	448	524	585	635	693	700	729	738	741	754	763
23	326	449	527	588	639	696	704	733	742	745	759	767
24	329	452	534	591	642	697	707	736	746	749	762	774

Table 8.2: (Continued)
Scaled Score to Percentile Rank Conversions by Grade (at month 7 in the school year)

	Grade											
PR	1	2	3	4	5	6	7	8	9	10	11	12
25	331	454	537	594	646	700	712	741	749	752	767	779
26	333	455	540	597	648	704	714	744	752	755	770	784
27	337	458	542	600	652	707	717	746	756	759	772	788
28	340	459	544	603	656	711	721	752	759	762	775	793
29	342	462	548	604	659	713	723	752	762	765	780	797
30	344	464	551	608	661	715	726	755	765	768	782	800
31	348	466	553	611	665	716	729	759	768	771	787	804
32	348	468	556	614	667	719	732	765	771	774	789	807
33	351	470	560	616	669	722	734	767	774	777	793	809
34	354	471	562	619	671	722	737	770	777	780	799	812
35	356	473	563	621	674	725	741	772	780	783	802	815
36	358	475	566	623	677	728	744	774	783	786	804	817
37	360	478	569	626	679	730	748	777	785	789	807	821
38	362	479	571	628	680	733	751	780	788	792	809	822
39	363	482	574	630	683	736	754	784	791	794	811	826
40	366	484	576	633	686	737	757	785	794	797	814	827
41	368	485	578	635	689	741	759	788	796	800	815	830
42	370	488	581	637	692	742	763	791	799	803	818	834
43	371	489	583	639	693	744	766	796	801	805	820	837
44	373	492	585	641	696	747	769	799	804	808	822	837
45	374	493	587	644	696	748	772	803	807	810	825	841
46	376	496	589	646	699	750	776	807	809	813	826	843
47	379	498	592	648	701	752	778	809	812	816	829	845
48	381	500	595	652	704	755	781	812	814	818	831	847

Table 8.2: (Continued)
Scaled Score to Percentile Rank Conversions by Grade (at month 7 in the school year)

	Grade											
PR	1	2	3	4	5	6	7	8	9	10	11	12
49	384	501	596	656	706	758	784	814	817	821	834	849
50	385	504	598	657	708	763	788	817	820	824	837	852
51	388	506	600	659	710	767	789	819	822	826	840	856
52	389	507	602	661	714	768	793	822	825	829	843	858
53	392	509	604	663	716	772	796	827	827	831	845	862
54	395	511	607	666	719	773	799	830	830	834	848	865
55	397	512	608	667	721	775	800	833	833	837	851	867
56	400	514	611	669	722	778	804	835	835	839	854	871
57	403	515	611	671	724	781	805	838	838	842	857	871
58	405	517	614	673	726	782	808	840	840	845	860	875
59	407	520	615	674	728	785	811	842	843	847	861	878
60	410	523	618	676	730	789	812	845	846	850	864	880
61	412	526	620	678	732	791	815	846	848	853	867	882
62	415	528	622	681	733	793	819	851	851	855	869	885
63	418	530	624	683	735	794	822	853	854	858	871	889
64	422	533	625	685	737	796	824	856	857	861	874	891
65	425	534	626	687	739	798	826	859	859	864	876	893
66	428	537	629	690	741	800	828	863	863	867	879	895
67	429	539	631	693	743	802	830	866	866	870	884	899
68	432	541	633	695	745	804	833	868	868	873	888	900
69	434	543	636	698	748	807	835	872	872	876	890	902
70	437	545	639	700	749	809	837	874	874	879	893	905
71	439	548	641	702	752	812	839	878	878	882	898	908
72	442	551	644	704	755	814	841	881	881	885	901	911

Table 8.2: (Continued)
Scaled Score to Percentile Rank Conversions by Grade (at month 7 in the school year)

	Grade											
PR	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
73	444	552	647	707	757	817	845	883	884	888	905	914
74	448	554	649	708	760	818	848	886	887	892	908	919
75	451	555	652	711	764	820	852	889	890	895	911	922
76	454	558	656	715	766	822	856	893	894	899	915	923
77	456	559	657	718	768	824	859	896	897	902	919	926
78	459	562	658	721	770	826	862	899	901	906	923	930
79	461	564	661	723	773	827	865	904	905	910	926	935
80	464	567	663	725	778	832	868	905	909	914	930	937
81	467	569	666	727	780	834	871	910	913	918	935	941
82	469	572	669	731	782	840	874	915	917	922	940	947
83	472	576	670	733	787	844	880	918	921	927	943	952
84	476	578	674	736	789	848	885	923	926	931	948	956
85	480	582	677	738	793	852	889	925	931	936	954	963
86	484	585	679	741	796	854	893	931	936	942	959	971
87	487	589	681	744	802	860	897	934	942	947	966	976
88	489	596	684	748	807	863	900	941	948	953	972	983
89	493	599	687	752	812	872	906	948	954	960	977	989
90	497	603	690	756	815	876	911	953	961	967	981	996
91	506	607	696	761	819	880	917	960	968	974	989	1006
92	513	614	700	765	826	890	923	968	977	983	997	1025
93	522	619	705	772	830	898	930	976	986	992	1006	1033
94	531	625	709	778	837	902	936	982	996	1003	1019	1042
95	539	632	718	786	842	909	943	992	1009	1015	1030	1060
96	551	644	727	795	853	919	956	998	1024	1030	1037	1077

Table 8.2: (Continued)
Scaled Score to Percentile Rank Conversions by Grade (at month 7 in the school year)

	Grade											
PR	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
97	567	653	736	808	863	931	967	1008	1030	1038	1054	1102
98	582	680	748	825	879	959	988	1030	1053	1061	1069	1112
99	612	759	780	848	897	1000	1029	1046	1087	1095	1108	1121

Table 8.3:
Percentile Rank to Normal Curve Equivalent Conversions

PR	NCE
1	1.0
2	6.7
3	10.4
4	13.1
5	15.4
6	17.3
7	18.9
8	20.4
9	21.8
10	23.0
11	24.2
12	25.3
13	26.3
14	27.2
15	28.2
16	29.1
17	29.9
18	30.7
19	31.5
20	32.3

Table 8.3: (Continued)
Percentile Rank to Normal Curve Equivalent Conversions

PR	NCE
21	33.0
22	33.7
23	34.4
24	35.1
25	35.8
26	36.5
27	37.1
28	37.7
29	38.3
30	39.0
31	39.6
32	40.1
33	40.7
34	41.3
35	41.9
36	42.5
37	43.0
38	43.6
39	44.1
40	44.7
41	45.2
42	45.8
43	46.3
44	46.8
45	47.4
46	47.9
47	48.4
48	48.9
49	49.5
50	50.0

Table 8.3: (Continued)
Percentile Rank to Normal Curve Equivalent Conversions

PR	NCE
51	50.5
52	51.1
53	51.6
54	52.1
55	52.6
56	53.2
57	53.7
58	54.2
59	54.8
60	55.3
61	55.9
62	56.4
63	57.0
64	57.5
65	58.1
66	58.7
67	59.3
68	59.9
69	60.4
70	61.0
71	61.7
72	62.3
73	62.9
74	63.5
75	64.2
76	64.9
77	65.6
78	66.3
79	67.0
80	67.7

Table 8.3: (Continued)
Percentile Rank to Normal Curve Equivalent Conversions

PR	NCE
81	68.5
82	69.3
83	70.1
84	70.9
85	71.8
86	72.8
87	73.7
88	74.7
89	75.8
90	77.0
91	78.2
92	79.6
93	81.1
94	82.7
95	84.6
96	86.9
97	89.6
98	93.3
99	99.0

Table 8.4:
Normal Curve Equivalent to Percentile Rank Conversions

NCE Range	PR
Low - High	
1.0 - 4.0	1
4.1 - 8.5	2
8.6 - 11.7	3
11.8 - 14.1	4
14.2 - 16.2	5
16.3 - 18.0	6
18.1 - 19.6	7
19.7 - 21.0	8
21.1 - 22.3	9
22.4 - 23.5	10
23.6 - 24.6	11
24.7 - 25.7	12
25.8 - 26.7	13
26.8 - 27.6	14
27.7 - 28.5	15
28.6 - 29.4	16
29.5 - 30.2	17
30.3 - 31.0	18
31.1 - 31.8	19
31.9 - 32.6	20
32.7 - 33.3	21
33.4 - 34.0	22
34.1 - 34.7	23
34.8 - 35.4	24
35.5 - 36.0	25
36.1 - 36.7	26
36.8 - 37.3	27
37.4 - 38.0	28
38.1 - 38.6	29

Table 8.4: (Continued)
Normal Curve Equivalent to Percentile Rank Conversions

NCE Range	PR
38.7 - 39.2	30
39.3 - 39.8	31
39.9 - 40.4	32
40.5 - 40.9	33
41.0 - 41.5	34
41.6 - 42.1	35
42.2 - 42.7	36
42.8 - 43.2	37
43.3 - 43.8	38
43.9 - 44.3	39
44.4 - 44.9	40
45.0 - 45.4	41
45.5 - 45.9	42
46.0 - 46.5	43
46.6 - 47.0	44
47.1 - 47.5	45
47.6 - 48.1	46
48.2 - 48.6	47
48.7 - 49.1	48
49.2 - 49.7	49
49.8 - 50.2	50
50.3 - 50.7	51
50.8 - 51.2	52
51.3 - 51.8	53
51.9 - 52.3	54
52.4 - 52.8	55
52.9 - 53.4	56
53.5 - 53.9	57
54.0 - 54.4	58
54.5 - 55.0	59

Table 8.4: (Continued)
Normal Curve Equivalent to Percentile Rank Conversions

NCE Range	PR
55.1 - 55.5	60
55.6 - 56.1	61
56.2 - 56.6	62
56.7 - 57.2	63
57.3 - 57.8	64
57.9 - 58.3	65
58.4 - 58.9	66
59.0 - 59.5	67
59.6 - 60.1	68
60.2 - 60.7	69
60.8 - 61.3	70
61.4 - 61.9	71
62.0 - 62.5	72
62.6 - 63.1	73
63.2 - 63.8	74
63.9 - 64.5	75
64.6 - 65.1	76
65.2 - 65.8	77
65.9 - 66.5	78
66.6 - 67.3	79
67.4 - 68.0	80
68.1 - 68.6	81
68.7 - 69.6	82
69.7 - 70.4	83
70.5 - 71.3	84
71.4 - 72.2	85
72.3 - 73.1	86
73.2 - 74.1	87
74.2 - 75.2	88
75.3 - 76.3	89

Table 8.4: (Continued)
Normal Curve Equivalent to Percentile Rank Conversions

NCE Range	PR
76.4 - 77.5	90
77.6 - 78.8	91
78.9 - 80.2	92
80.3 - 81.7	93
81.8 - 83.5	94
83.6 - 85.5	95
85.6 - 88.0	96
88.1 - 91.0	97
91.1 - 95.4	98
95.5 - 99.0	99

Chapter 9: Growth Measurement

New interventions are continually being proposed for educational settings, most with the aim of improving educational outcomes. Such interventions may be extensive, such as a new teaching method or new curriculum, or they may be smaller in scope, such as a new textbook. The introduction of a learning information system (LIS), such as Accelerated Math, into a school or classroom is a good example of such an intervention. Whatever the proposed intervention, however, it is first necessary to establish its effectiveness in terms of the educational benefit for students. Examination of the effectiveness of new teaching methods, a new curriculum, and other such interventions is extremely important to accurately determine whether these programs and/or methods indeed help students. This examination, often called growth measurement, is important for the appropriate direction of limited educational funds and for ensuring that programs that will have the most educational impact on children are clearly identified.

Absolute Growth and Relative Growth

It is important to distinguish between two types of academic growth (or gains) that may be evidenced in test results: absolute growth and relative growth. Absolute growth reflects any and all growth that has occurred. On the other hand, relative growth reflects only growth that is above and beyond "normal" growth (i.e., beyond typical growth in a reference or norming group).

As an example, imagine a group of students whose test results place them at the 40th percentile, with an average Scaled Score of 686, in the spring of grade 5. In the spring of grade 6, the same group still scores at the 40th percentile with an average Scaled Score of 737. This group of students has experienced 51 Scaled Score points of absolute growth, but there has been no relative growth (since the group scored at the 40th percentile in both grade distributions). In other words, relative growth will only be positive when growth has exceeded "normal" growth as defined by the norming sample. In general, norm-referenced scores, such as Percentile Ranks, only indicate relative growth, whereas Scaled Scores (and Grade Equivalent scores) reflect absolute growth.

The Growth Report in STAR Math 3.x provides you with information about both aspects of growth. In general, most educational program evaluation designs attempt to determine if relative growth has occurred. That is, they attempt to measure the impact of the intervention, or program, above and beyond normal growth (i.e., above and beyond what you would expect to occur without the intervention).

The Pretest/Posttest Paradigm for Measuring Growth

One popular method for measuring growth (i.e., measuring effectiveness of educational interventions) includes the use of a pretest/posttest design. In such a design, each student is administered a test prior to the beginning of the intervention to establish a baseline measure or a measure of the student's achievement level before the intervention. Then, each student is measured again at a later point in time (usually with a different, but equated, "form" of the same test) to see if the intervention is providing the desired outcome. The follow-up measurement may be at the end of the intervention, or it may be done periodically throughout the course of the new program.

The psychometric characteristics of the test, such as its reliability and validity, must be strong in order for this type of research to work properly. Additionally, if the same test is used for both the pretest and the posttest, the results may be compromised due to students having previously been exposed to the test items. Ideally, equivalent, or "parallel," tests with no items in common should be administered. Therefore, subsequent administration of a computer-adaptive test, such as STAR Math 3.x, is ideal for the pretest-posttest method because its proprietary software will generate psychometrically parallel versions of the test at both pretest and posttest administrations, with no common items.

It is important to note that growth is best measured at a group level, such as a classroom or grade level. This is because at the individual student level there are technical issues of unreliability associated with growth (gain) scores.

Pretest/Posttest with Control Group Design

In the "classic" implementation of a pretest/posttest design, the group (classroom, grade, or school) receiving the new intervention is referred to as the experimental group. A second, matched group that does not receive the intervention is referred to as the control group. The control group follows the same pre- and posttesting pattern in order to serve as a baseline for "normal" growth without the intervention. Growth is indicated when the difference between the groups' average (mean) scores (computed as posttest mean score minus pretest mean score) is positive. Because it is likely that growth will occur even if the program (or intervention) is ineffective, the program's effectiveness (or impact) is measured when the growth for the experimental group is significantly greater than growth for the control group.

Pretest/Posttest without a Control Group Design

When the test scores used in the evaluation are norm-referenced, such as Percentile Ranks, a control group is not necessarily required because the scores themselves allow the measurement of growth relative to the peer (norming) group. Because this method only requires a single group, many researchers use this type of design. Furthermore, without a control group all students may receive the intervention.

It should be noted that when a test is normed, the Percentile Rank information is derived based on the specific point during the academic year when the test was administered. For example, assume that a test was normed in the spring (7 months into the school year), but a teacher wants to make an assessment at the beginning of the school year. To provide normative information for each month of the academic year, the difference between adjacent grade levels is examined and, presuming even growth, values are generated using interpolation. Because the assumption that growth occurs evenly over an academic year may be unrealistic, teachers should use caution when looking at growth based on these interpolated percentiles.

Using Scores to Measure Growth

There are a number of pieces of score information that are available from a standardized test such as STAR Math. Among the scores available from the test are Scaled Scores, Percentile Ranks, Normal Curve Equivalents, and Grade Equivalents. All of these scores appear on the STAR Math 3.x Growth Report, and all may be useful in certain situations for examining growth. As described in the following sections, the selection of the specific score(s) for measuring growth depends on the test user's goals.

Scaled Scores

Scaled Scores represent the student's score as expressed on a continuous vertical scale that spans all grade levels (1-12). The underlying vertical scale is derived as part of the test development process. In adaptive testing, students can receive different sets of items and still receive a comparable Scaled Score that represents their unique underlying ability level. Because Scaled Scores essentially map a student to a specific location on the underlying ability continuum, they can be useful in measuring absolute growth, and they are included in the Growth Report in STAR Math 3.x.

Percentile Ranks

Percentile Ranks provide an easy way of relating a student's Scaled Score to the performance of a specified norming group (i.e., relating performance to one's peers). By providing a reference to a "standard" (i.e., norming group), norm-referencing adds a different meaning to scores than Scaled Scores. Percentile Ranks range from 1 through 99 and define the percent of the norming sample that achieved lower Scaled Scores.

Table 8.2 (page 80) lists the conversions from STAR Math 2.0/3.0 Scaled Scores to Percentile Ranks. In this *Technical Manual*, the published Scaled Score to Percentile Rank conversion tables are abridged, showing only those values that are appropriate for administration of a test in the seventh month of the school year. However, within the STAR Math 3.x software, the appropriate point-in-time Scaled Score to Percentile Rank relationship is determined dynamically as a normal part of the scoring process.

Because they are relatively easy to understand and to explain to others, Percentile Ranks are probably the most common method for expressing results on norm-referenced tests. For example, if a Scaled Score of 648 corresponds to a Percentile Rank of 47 (at grade 4), it means that 47 percent of the norming group had Scaled Scores lower than 648. However, the main disadvantage of using Percentile Ranks is that they are not on an equal interval scale in terms of the underlying skill level. Quite simply, this means that gains of 1 percentile point, at various points along the scale, do not represent equal gains in achievement (skills). Because equal units on the percentile scale do not represent equal amounts of the underlying ability, it is not appropriate or meaningful to compute averages based on Percentile Ranks.

Normal Curve Equivalents

Before Percentile Ranks can be averaged, they must undergo a transformation that places them onto a scale that does have equal interval properties, usually the Normal Curve Equivalent (NCE). Although the STAR Math 3.x software automatically performs this score transformation in the Growth Report, the tables used for converting Percentile Ranks to Normal Curve Equivalents (and from NCE to PR) are also provided in Tables 8.3 (page 84) and 8.4 (page 88).

To obtain a measure of average performance in terms of Percentile Ranks, it is important to first transform the Percentile Ranks to an equal interval scale by converting them into NCE values. Then, the mean or average of these NCEs can be calculated. This is the same process used internally for reporting group Percentile Rank information in STAR Math 3.x. Once the NCE scores have been averaged, it is allowable to "map" the resulting mean NCE back to its corresponding Percentile Rank for reporting of growth.

Grade Equivalent Scores

Grade Equivalent (GE) scores represent one of the most commonly used scores for comparisons with a norms group. It is very important, however, to note that the GE scale is only loosely tied to the empirical normative data, specifically only at the 50th percentile points at each grade level. GE scores are popular because they appear to be easy to understand. However, GE scores can often be more misleading than helpful and are quite commonly misinterpreted. For example, if a sixth-grade student scores at Grade Equivalent level 4.1 on a math test, that does not mean that the sixth-grader is only capable of fourth-grade work. Technically, it means that the sixth-grader achieved a score that would be comparable with an "average," or at the 50th percentile, fourth-grade student after one month of instruction if they had taken the same test that the sixth-grader received.

In an adaptive test, such as STAR Math, the branching capabilities may result in a student completing a test that is similar in content and difficulty to what students at the grade-appropriate level would have received, and this often leads to the misinterpretation of GEs. It is important to note, however, that the GE score provides only a rough estimate of the student's achievement level.

Most teachers would agree that the amount of growth in terms of math skill is not the same between, for example, grades 1 and 2 as between grades 9 and 10. In fact, it is easily verifiable that the gains in math, in terms of the underlying ability, are much greater in the early grades and tend to diminish with increasing grade level.

For norm-referenced tests, such as STAR Math, the rate of growth on the GE scale within any particular grade has been estimated based on a relatively small number of points, corresponding to times at which the testing actually occurred. As noted above, the curve that joins these points is usually interpolated by assuming that the rate of growth (or progress between these points) is even and continuous. In a subject such as math, for example, this assumption of equal amounts of progress from month to month throughout the school year may be untenable. To demonstrate, if two students have GE scores of 3.8 and 3.9, respectively, it is not necessarily true that giving an additional month of instruction to the first student would place them at the same functional level.

GEs should not be used as the standard for growth per year or per grade because "one year's growth," as measured in GE scores, varies across the student distribution. During a one-year period, the "normal" growth in GE scores would be 1.0 *only at the 50th percentile*. In general, below the 50th percentile, less than one year's growth (in terms of GEs) occurs during a one-year period, and above the 50th percentile, more than one year's growth (in terms of GEs) occurs during a one-year period.

For example, using the STAR Math 2.0 norms, a student at the 15th percentile at grade 4 (Scaled Score of 559) had a GE of 3.2. If the student achieved "normal" growth and remains at the 15th percentile in grade 5 (Scaled Score of 607), her GE would be 3.8, an increase of only +0.6 GE units. A student at the 85th percentile in grade 4 (Scaled Score of 738) has a GE of 6.2. If this student achieved "normal" growth and remained at the 85th percentile in grade 5 (Scaled Score of 793), his GE would be 7.8, an increase of +1.6 GE units.

As with Percentile Ranks, GE scores are not on an equal-interval scale and, therefore, should not be averaged in order to obtain the "typical" GE for the group. GEs on the Growth Report and the Summary Report in STAR Math 3.x correspond to the average of the students' Scaled Scores rather than averaging Grade Equivalents. It is important to note that this "typical" Grade Equivalent is properly interpreted as the Grade Equivalent that corresponds to average student performance, instead of interpreting it as the average of the Grade Equivalents. Although it may be a subtle difference, this interpretation is not the same as the average GE, and the two methods will not necessarily yield the same computational results.

Pretest/Posttest Studies of Growth Using a Single Sample Referenced Against Normative Data

The goal of this type of study is to determine if a program intervention has resulted in improvement beyond what is expected based on the norming population (i.e., to see if the posttest results place the students above where they would be if there had not been any intervention). For example, if a group of 4th-grade students' pretest scores indicate that their group percentile (corresponding to the average NCE) is 25, then we want to see if their 5th-grade posttest scores will result in a group percentile that is greater than 25.

When comparing these students' growth to growth based on norms, only one group is required, but in this case, the time period between pretest and posttest should be at least one year; otherwise the growth would be referenced against interpolated data. This corresponds with U. S. Department of Education recommendations for Chapter I (Title I) program impact studies, which state that:

> *The general rule of thumb for norm-referenced evaluations is that testing should be done within two weeks of the midpoint of the empirical norming period (U. S. D. E. Evaluator's References for Title I Evaluation and Reporting System, Volume 2).*

For STAR Math 3.x, the empirical norming period was in the seventh month of the school year. The U. S. Department of Education further recommends that interpolated norms that vary by more than six weeks from the empirical data points should not be used for norm-referenced evaluations.

In general, a good guideline regarding sample size requirements for any growth study is "more is better." As the size of the group increases, you can be more confident that the obtained results are genuine.

STAR Math 3.x and the Elementary and Secondary Educational Act (ESEA, No Child Left Behind)

STAR Math may be useful for districts and schools as they conform to the 2002 No Child Left Behind Legislation. For example, according to "No Child Left Behind," starting in 2005, states must annually measure the math progress of students in grades 3-8. As noted throughout this manual, STAR Math is a reliable and valid measure of math achievement for students in grade 1-12. Furthermore, due to its computer-adaptive features, STAR Math requires less administration time and supervision than paper-and-pencil tests without compromising the psychometric quality of scores. No Child Left Behind also requires that federal funding go only to those math programs that are backed by scientific evidence. As noted in the above section on growth measurement, teachers and administrators can use STAR Math to evaluate the effectiveness of math programs and interventions. Given the increased emphasis being placed on using only research-based teaching methods, more and more teachers will find STAR Math an invaluable tool in the process of demonstrating growth in math achievement resulting from their math programs.

Chapter 10: Frequently Asked Questions

The STAR Math computer-adaptive test is designed to be user-friendly. However, because the topics of psychometrics and standardized assessment are quite complex, this section answers questions commonly asked about STAR Math.

What is the primary purpose of the STAR Math assessment? Why have so many schools purchased it, and how are they using the results?

STAR Math tests serve the same purposes as the highly recognized STAR Reading tests, only in a different content area. The STAR Math software allows teachers to:

- Place new students in the appropriate level of math instructional materials, or in the appropriate Accelerated Math library.
- Measure growth in math skills or the effectiveness of a math intervention program like the adoption of Accelerated Math throughout the school year.
- Predict how students will do on high-stakes tests while there is still time to intervene.

Because the STAR Math computer-adaptive math test is the only classroom-based assessment that can give teachers this kind of information in just 15 minutes, many educators find STAR Math an invaluable tool.

How can STAR Math accurately determine a student's math level with only 24 test questions and in just 15 minutes?

A low number of test questions and a short test time are possible because of STAR Math's advanced computer-adaptive technology. Adaptive Branching allows the test to very quickly adapt to the student's level of proficiency. The STAR Math program acquires new information about the student's math ability with each and every item and updates its knowledge of the student's ability after every question. This means that STAR Math tests are much more efficient than conventional paper-and-pencil tests that administer the same items regardless of how the student is doing. By obtaining more information from every item administered, and by using that information to continuously tailor items for the student, STAR Math tests are able to achieve measurement precision comparable to much longer conventional tests. This results in an efficient and reliable assessment for teachers and a positive testing experience for students.

What evidence do we have that STAR Math performs as claimed?

Evidence of STAR Math's performance is gathered in two forms: reliability and validity.

- Reliability is the extent to which a test yields consistent results from one administration to another and from one test form to another. Internal research studies suggest that STAR Math 2.x/3.x test scores have a very high level of internal consistency reliability, as well as a high degree of alternate-form reliability.
- Validity is the degree to which a test measures what it claims to measure. STAR Math 2.x/3.x test score validity is evidenced by the high correlation to overall math scores on many high-stakes standardized tests, as well as the high correlation between STAR Math 2.x/3.x Scaled Scores and teachers' ratings of their students' math skills.

See Chapter 6 (page 48) for more information on STAR Math 2.x/3.x reliability, and see Chapter 7 (page 54) for information on its validity.

What areas of the math curriculum are covered on STAR Math tests, and how were they selected?

The content of the STAR Math 2.x/3.x item bank reflects the majority of material covered in popular math textbooks, as well as state and national standards. Content objectives were chosen through an extensive review of leading math textbook series, state curriculum guidelines, National Council of Teachers of Mathematics (NCTM) standards, and specifications from the National Assessment of Educational Progress (NAEP) and the Trends in International Mathematics and Science Study (TIMSS.) As a result, STAR Math 2.x/3.x questions reflect the content most relevant to teachers and students. For more information on the specific content of the STAR Math 2.x/3.x item bank, refer to Chapter 2 (page 13).

There don't seem to be any calculus items. What are the most difficult questions in the test?

Because most of the items at the top of the difficulty scale are from the Geometry and Numeration Concepts (e.g. fractional exponents) content objectives, the STAR Math software may administer items from these strands to very high-performing students. The following features of the STAR Math test should also be noted:

- Algebra items are limited to the last section of the test. Content balancing considerations limit the number of algebra items administered during any test. At the highest grades and performance levels, at least two but no more than three algebra items will be administered. At lower grades and performance levels, algebra items will seldom be administered.

- Calculus items were not included on the test because typical high school students, both in the national norming sample and in the U.S. population as a whole, have not taken calculus. However, many items in the Algebra strand address material covered in many high school Pre-Calculus courses.

When I take a STAR Math test, I keep getting difficult questions even though I entered myself as a lower grade student. Why?

You are probably answering items correctly that a student at that grade level would normally get wrong. The grade you select for yourself only affects the difficulty of the first item on your first test. After that, the adaptive brancher takes over based on your responses. Subsequent tests begin just below your previously tested ability, regardless of your grade level.

To simulate the experience of a lower grade student, you would need to answer several questions incorrectly. (Alternating correct and incorrect responses will approximately maintain the difficulty level, while more correct or incorrect answers will cause it to move up or down the difficulty scale, respectively.) Because it is quite difficult for most adults to "act like" young students when completing a STAR Math test, teachers wishing to evaluate the software should observe an actual administration with a student.

There doesn't seem to be any pattern to the types of STAR Math test questions posed. How does it select the math objectives to be tested on?

All STAR Math 3.x tests follow a similar pattern: the first eight STAR Math 3.x items measure Numeration Concepts, the 9^{th} through the 16^{th} items measure Computation Processes, and the last eight items measure other applications in six strands of math objectives.

During a STAR Math test, items are also selected so that they are the appropriate difficulty for each student. All of the questions in the item bank, from all math content and objective areas, were placed on the same difficulty scale through a process called calibration. During a STAR Math 3.x test, the adaptive brancher moves up and down that difficulty scale, selecting the next item based on the student's current ability estimate. Item selection is based primarily on the calibrated difficulty of the questions.

Finally, steps are also taken to ensure a variety of objectives are assessed. The probability of receiving an item from a specific topic area or objective depends largely on the concentration of such items in the pool around the estimated ability level on the difficulty scale.

See Chapter 2 (page 13) for information about the content strands and objectives, and refer to Appendix A (page 106) for a list of the STAR Math 2.x/3.x objectives.

My students get items on material we haven't covered yet. Can this be prevented?

Not entirely. This is the nature of computer-adaptive testing, the technology that permits you to get accurate test results in only 15 minutes. If a student is performing well, the STAR Math software continues to administer more difficult items until it finds a level at which the student cannot answer questions correctly. Just as STAR Reading tests may "branch up" to vocabulary the student has not been exposed to, STAR Math tests may move up to content objectives a student hasn't yet reached. However, to minimize this phenomenon, the STAR Math software will not administer items that are four or more grade levels above the student's specified grade level. In addition, because, on average, students answer about 75 percent of STAR Math 3.x items correctly, students should not receive items on unfamiliar content frequently within a test.

The STAR Math test seems too difficult and frustrating for my higher-performing elementary school students.

The adaptive brancher is set so that, on average, students will answer about three fourths (75%) of the test items correctly. High-performing students in particular may be accustomed to getting much higher percentages correct on tests. These students should be instructed to expect a difficult test and to do their best without worrying about the number of correct or incorrect items.

May students use calculators or reference materials during a STAR Math test?

No. STAR Math tests are standardized. In order for the normative scores to be meaningful, STAR Math must be administered in the same way it was during the norming study. During the norming study, students were allowed to use blank scratch paper and a pencil, but not calculators or any reference materials. All STAR Math 3.x kits include *Pretest Instructions* that teachers can also use to make sure that the test administration is standardized. Because any variance from these procedures could invalidate students' scores, teachers should closely follow these instructions.

Why does the STAR Math software allow us to enter students in Kindergarten if they cannot be tested?

The student records are part of a multi-product central database that is accessible by any Renaissance Place Edition software program. Because students in grades K-12 may be entered in one program (e.g. STAR Math software) and then enrolled in another (e.g. STAR Early Literacy software), Kindergarten students may be enrolled in a class in STAR Math 3.x even though they cannot be tested.

Does the STAR Math test assess problem-solving or critical thinking skills?

Yes. The STAR Math item bank includes a Word Problems strand that closely parallels the Computation Processes strand. These word problems ensure that students can perform simple situational analyses. More difficult word problems also require a second computation or include extraneous information.

Why did you choose to use multiple-choice questions to measure problem-solving skills rather than open-ended questions?

The STAR Math test is designed to gather the maximum amount of information on problem solving and other math skills and to provide Percentile Rank and Grade Equivalent scores in the shortest period of time. Only multiple-choice type questions fit this purpose. Open-ended questions are more appropriate for teachers to use in a classroom setting when diagnosing any difficulties a particular student might be having.

How often should we administer STAR Math tests?

STAR Math tests may be administered up to five times a year. However, we recommend giving the test three times: near the beginning, middle, and end of the school year. Growth over shorter time periods may be masked by the standard error of the test (which is actually quite low and similar to much longer paper-and-pencil standardized tests). New students, or students for whom you occasionally need additional information, may be tested at any time. However, when measuring growth over time periods of less than one year, it is best to look at whole class, grade, or school averages rather than individual test scores. Scores for larger groups have lower standard errors and, because these scores are more reliable, teachers will more easily understand them.

What is the difference between Percentile Rank (PR) and Normal Curve Equivalent (NCE) scores?

PR and NCE scores are both presented on scales that range from 1 to 99 with a mean of 50. However, the increments between 1 and 50 or 50 and 99 are very different on the two scales.

- PR scores express performance relative to the performance of the students in the norms group. For example, if a student's PR is 50, his score is higher than 50 percent of the scores obtained by students at the same grade level when the test was normed. While they are useful for score interpretation, PR scores are not generally considered to have scale properties that are appropriate for statistical calculations. For this reason, mathematical operations, such as averaging scores for a group, should generally not be done using PR scores.

- NCE scores are derived from PRs by transforming them to an equal-interval scale. NCE scores may be averaged and then converted back to PRs. Because they can be mathematically manipulated and readily compared, NCE scores are often required for district and state reporting. Unlike PR scores, however, NCE scores are not readily interpretable. Therefore, many educators prefer to convert averaged NCE values back to PRs for interpretation.

Some of my students had their Grade Equivalent (GE) score go up but their Percentile Rank (PR) score go down. How can this be?

GE and PR scores measure two very different types of growth: "absolute growth" and "relative growth," respectively.

- Absolute growth reflects any and all growth that has occurred over a period of time, and is reported in terms of the Scaled Score (SS) and GE in STAR Math reports.
- Relative growth reflects growth relative to a peer group (such as the norms sample), and is reported in terms of PR and NCE scores in STAR Math reports.

Therefore, because they measure different types of growth, GE and PR scores may not increase and decrease together. For example, when a student's Grade Equivalent increases less than that of the student's peers, he will lose ground in terms of Percentile Rank. When this occurs, students' STAR Math Growth Reports will display increases in GE scores but decreases in PR scores. Refer to Chapter 9 (page 92) for more information on absolute and relative growth.

Several of our 8th graders are getting GE scores of 12.9+ on STAR Math tests. They aren't doing high school work, so how can they be scoring so high?

A GE of 12.9+ does not necessarily mean your students are capable of high school work. It only means that their Scaled Scores were at the same level as typical (50th percentile) high school seniors.

Because some high school students complete very advanced math courses, such as calculus, while others take no math courses after their first year or two of high school, there is significant variation in math abilities among high school students. This phenomenon is reflected in the STAR Math 2.0 norming sample's score distributions: the variability of scores at junior high and high school grades is larger than the range of average scores between, for example, grades 8 and 12. This means, inevitably, some students will obtain GE scores much higher than their own grade placements.

It should also be noted that GE scores can be misleading. A GE score of 12.9 denotes that the student's STAR Math Scaled Score was equal to or exceeded the 50^{th} percentile of students in the 9^{th} month of 12^{th} grade. However, it does NOT denote anything about students' exposure to specific topics in math, nor about their proficiency in specific objectives in which they have not yet received instruction.

For that reason, most educators find GE math scores most meaningful in the lower grades where all students go through a similar math curriculum and where grade-to-grade growth is more substantial. Secondary grade teachers will find STAR Math Scaled Scores (SS) and Percentile Rank (PR) scores more helpful than GEs in measuring ongoing math achievement.

Are STAR Math test results really very useful at the high school level?

Yes. STAR Math tests measure a wide range of math abilities at the high school level. Scaled Scores range from about 500 to 1200 for 12th graders, with 852 being the 50th percentile. For students beyond the Grade Equivalent scale (51st percentile for 12th graders by definition), STAR Math continues to do an excellent job of measuring "absolute" growth by means of the Scaled Score scale and "relative" growth by means of the Percentile Rank scale. The STAR Math test also does a very good job of measuring the math skills of incoming students and therefore helps high school math teachers quickly assess how prepared new students are for their math classes.

Is there a way for the teacher to see which questions a student answered correctly and incorrectly?

No. This is prevented for the following two reasons. First, in computer-adaptive normative testing, the student's performance on individual items is not as meaningful as the pattern of his or her responses on the entire test. The student's pattern of performance on all items taken together forms the basis of the scores in STAR Math reports. Second, for purposes of test security, preventing item review protects the test items from compromise and overexposure.

Explain what "calibration" and "norming" mean.

Development of the STAR Math 2.x/3.x normative assessment required two major phases of student testing: calibration and norming.

- Calibration is the process of placing individual test items on a difficulty scale. Calibration occurs by having a large number of students test on all of the questions to be included in the item bank and analyzing the resulting item response data. The difficulty scale is then used by STAR Math software for item selection using the Adaptive Branching algorithm, and to estimate the student's math ability level.
- Norming is the process of determining how a nationally representative sample of students at each grade level performs on the overall test. For STAR Math 2.0, a large number of students from grades 1 through 12 were tested using the final computer-adaptive test. An analysis of their ability estimates was then conducted in order to derive the GE and PR scoring tables.

See Chapter 3 (page 27) for information about the STAR Math 2.0 Calibration Study, and see Chapter 4 (page 36) for information about the STAR Math 2.0 Norming Study.

Appendix A: Strands and Objectives

Numeration Concepts

NA1	Ones: Placing numerals in order
NA2	Ones: Using numerals to indicate quantity
NA3	Ones: Relate numerals and number words
NA4	Ones: Use ordinal numbers
N00	Ones: Locate numbers on a number line
N01	Tens: Place numerals (10-99) in order of value
N02	Tens: Associate numeral with group of objects
N03	Tens: Relate numeral and number word
N04	Tens: Identify one more/one less across decades
N05	Tens: Understand the concept of zero
N06	Hundreds: Place numerals in order of value
N07	Hundreds: Relate numeral and number word
N08	Hundreds: Identify place value of digits
N09	Hundreds: Write numerals in expanded form
N11	Thousands: Place numerals in order of value
N12	Thousands: Relate numeral and number word
N13	Thousands: Identify place value of digits
N14	Thousands: Write numerals in expanded form
N16	Ten thousands, hundred thousands, millions, billions: Place numerals in order of value
N17	Ten thousands, hundred thousands, millions, billions: Relate numeral and number word
N18	Ten thousands, hundred thousands, millions, billions: Identify place value of digits
N19	Ten thousands, hundred thousands, millions, billions: Write numerals in expanded form
N21	Fractions and decimals: Convert fraction to equivalent fraction
N22	Fractions and decimals: Convert fraction to decimal
N23	Fractions and decimals: Convert decimal to fraction

Numeration Concepts (Continued)

N24	Fractions and decimals: Read word names for decimals to thousandths
N25	Fractions and decimals: Identify place value of digits in decimals
N26	Fractions and decimals: Identify position of decimals on number line
N27	Fractions and decimals: Identify position of fractions on number line
N28	Fractions and decimals: Convert improper fraction to mixed number
N29	Fractions and decimals: Round decimals to tenths, hundredths
N30	Fractions and decimals: Relate decimals to percents
N31	Advanced concepts: Determine square roots of perfect squares
N32	Advanced concepts: Give approximate square roots of a number
N33	Advanced concepts: Recognize meaning of nth root
N34	Advanced concepts: Recognize meaning of exponents (2-10)
N35	Advanced concepts: Recognize meaning of negative exponents
N36	Advanced concepts: Recognize meaning of fractional exponents
N37	Advanced concepts: Can use scientific notation
N38	Advanced concepts: Knows meaning of primes and composites
N39	Advanced concepts: Can determine greatest common factor
N40	Advanced concepts: Can determine least common multiple
N41	Advanced concepts: Recognizes use of negative numbers

Computation Processes

C01	Addition of basic facts to 10
C02	Subtraction of basic facts to 10
C03	Addition of basic facts to 18
C04	Subtraction of basic facts to 18
C05	Addition of three single digit addends
C06	Addition beyond basic facts, no regrouping (2d+1d)
C07	Subtraction beyond basic facts, no regrouping (2d-1d)

Computation Processes (Continued)

C08	Addition beyond basic facts with regrouping (2d+1d, 2d+2d)
C09	Subtraction beyond basic facts with regrouping (2d-1d, 2d-2d)
C10	Addition beyond basic facts with double regrouping (3d+2d, 3d+3d)
C11	Subtraction beyond basic facts with double regrouping (3d-2d, 3d-3d)
C12	Multiplication basic facts
C13	Division basic facts
C14	Multiplication beyond basic facts, no regrouping (2dx1d)
C15	Division beyond basic facts, no remainders (2d/1d)
C16	Multiplication with regrouping (2dx1d, 2dx2d)
C17	Division with remainders (2d/1d, 3d/1d)
C18	Addition of whole numbers: any difficulty
C19	Subtraction whole numbers: any difficulty
C21	Division of whole numbers: any difficulty
C22	Addition of fractions: like single digit denominators
C23	Subtraction of fractions: like single digit denominators
C24	Addition of fractions: unlike single digit denominators
C25	Subtraction of fractions: unlike single digit denominators
C26	Multiplication of fractions: single digit denominators
C27	Division of fractions: single digit denominators
C28	Addition of mixed numbers
C29	Subtraction of mixed numbers
C30	Multiplication of mixed numbers
C31	Division of mixed numbers
C33	Addition of decimals, place change (e.g. 2 + .45)
C35	Subtraction of decimals, place change (e.g. 5-.4)
C36	Multiplication of decimals
C37	Division of decimals

Computation Processes (Continued)

C38	Percent A (10 is what % of 40)
C39	Percent B (20% of 50 is what)
C40	Percent C (30 is 50% of what)
C41	Proportions
C42	Ratios

Other Applications

Estimation	
E06	Estimation problems: Addition beyond basic facts, no regrouping (2d+1d)
E07	Estimation problems: Subtraction beyond basic facts, no regrouping (2d-1d)
E14	Estimation problems: Multiplication beyond basic facts, no regrouping (2dx1d)
E15	Estimation problems: Division beyond basic facts, no remainders (2d/1d)
E18	Estimation problems: Addition of whole numbers, any difficulty
E19	Estimation problems: Subtraction of whole numbers, any difficulty
E20	Estimation problems: Multiplication of whole numbers, any difficulty
E21	Estimation problems: Division of whole numbers, any difficulty
E24	Estimation problems: Addition of fractions, unlike single digit denominators
E25	Estimation problems: Subtraction of fractions, unlike single digit denominators
E28	Estimation problems: Addition of mixed numbers
E29	Estimation problems: Subtraction of mixed numbers
E32	Estimation problems: Addition of decimals, no place change (e.g. 2.34+10.32)
E33	Estimation problems: Addition of decimals, place change (e.g. 2+.45)
E34	Estimation problems: Subtraction of decimals, no place change (e.g. .53 - .42)
E35	Estimation problems: Addition of decimals, place change (e.g. 5+.4)
E38	Estimation problems: Percent A (10 is what % of 40)
E39	Estimation problems: Percent B (20% of 50 is what)
E40	Estimation problems: Percent C (30 is 50% of what)

Other Applications (Continued)

Geometry	
GA1	Use basic terms to describe position
GA2	Identify common plane shapes
GA3	Identify common plane shapes when rotated
GA4	Compare common objects to basic shapes
GA5	Understand basic symmetry
GA6	Recognize elements of basic shapes
GA7	Identify common solid shapes
G00	Identify fraction parts of common plane shapes
G01	Identify numeric patterns
G02	Circle terms
G03	Perimeter: square
G04	Perimeter: rectangle
G05	Perimeter: triangle
G06	Area: square
G07	Area: rectangle
G08	Area: right triangle
G09	Area: circle
G10	Volume: rectangular prism
G12	Identify rays
G13	Identify line segments
G14	Identify parallel lines
G15	Identify intersecting lines
G16	Identify perpendicular lines
G17	Use properties of parallel lines
G18	Use properties of intersecting lines
G19	Use properties of perpendicular lines

Other Applications (Continued)

G20	Vertical and supplementary angles
G21	Classify angles (obtuse, etc.)
G22	Using parts of a triangle
G23	Pythagorean theorem
Measurement	
MA1	Use simple vocabulary of measurement
MA2	Understand the value of penny, nickel, dime
MA3	Understand the value of groups of coins to $1.00
MA4	Determine the value of quarter and dollar
MA5	Tell time to the hour and half hour
MA6	Read a thermometer
MA7	Order days of the week
M00	Order months of the year
M01	Customary measures: Inches, feet, yards
M02	Customary measures: Estimating linear measures
M03	Customary measures: Estimating volume measures
M04	Customary measures: Pints, quarts, gallons
M05	Metric prefixes
M06	Metric: Customary conversions
M07	Measures of angles
M08	Estimating linear measure in metric units.
Data Analysis and Statistics	
SA1	Read tally charts
S00	Read simple pictographs
S01	Read table
S02	Read bar graph
S03	Read pie graph

Other Applications (Continued)

S04	Interpret table
S05	Interpret bar graph
S06	Interpret pie graph
S07	Statistics: Mean
S08	Statistics: Median
S11	Probability: Simple
S12	Probability: Joint
Word Problems	
W03	Word problems: Addition of basic facts
W04	Word problems: Subtraction of basic facts
W06	Word problems: Addition beyond basic facts, no regrouping (2d+1d)
W08	Word problems: Addition beyond basic facts with regrouping (2d+1d, 2d+2d)
W09	Word problems: Subtraction beyond basic facts with regrouping (2d-1d, 2d-2d)
W12	Word problems: Multiplication of basic facts
W13	Word problems: Division of basic facts
W14	Word problems: Multiplication beyond basic facts, no regrouping (2dx1d)
W15	Word problems: Division beyond basic facts, no remainders (2d/1d)
W16	Word problems: Multiplication with regrouping (2dx1d, 2dx2d)
W17	Word problems: Division with remainders (2d/1d, 3d/1d)
W18	Word problems: Addition of whole numbers, any difficulty
W19	Word problems: Subtraction of whole numbers, any difficulty
W20	Word problems: Multiplication of whole numbers, any difficulty
W21	Word problems: Division of whole numbers, any difficulty
W22	Word problems: Addition of fractions, like single digit denominators
W23	Word problems: Subtraction of fractions, like single digit denominators
W24	Word problems: Addition of fractions, unlike single digit denominators
W25	Word problems: Subtraction of fractions, unlike single digit denominators

Other Applications (Continued)

W28	Word problems: Addition of mixed numbers
W29	Word problems: Subtraction of mixed numbers
W2S	Word problems: Two-step
W33	Word problems: Addition of decimals, place change (e.g. 2+.45)
W35	Word problems: Subtraction of decimals, place change (e.g. 5-.4)
W36	Word problems: Multiplication of decimals
W37	Word problems: Division decimals
W38	Word problems: Percent A (10 is what % of 40)
W39	Word problems: Percent B (20% of 50 is what)
W40	Word problems: Percent C (30 is 50% of what)
W41	Word problems: Proportions
W42	Word problems: Ratios
WXI	Word problems: Extra information
Algebra	
A00	Can skip count by 2, 5, 10 in ascending order
A01	Simple number sentence
A02	Translate word problem to equation
A03	Linear equations: 1 unknown
A04	Linear equations: 2 unknowns
A05	Reciprocals of rational numbers
A06	Graph of linear equation (integers add, subtract)
A07	Linear inequalities: 1 unknown
A08	Linear inequalities: 2 unknown
A09	Graph linear inequalities
A10	Classify mono, bi, or trinomials
A11	Polynomials: Order polynomials
A12	Polynomials: Addition and subtraction

Other Applications (Continued)

A13	Polynomials: Multiplication and division
A14	Solve system of 2 equations (2 unknowns)
A15	Quadratic equations: Solve using square root rule
A16	Quadratic equations: Solve by factoring
A17	Quadratic equations: Completing the square
A18	Factor common term from binomial expression
A19	Determine slope
A20	Determine intercept
A21	Sequences and series: Common differences in arithmetic sequences
A22	Sequences and series: Find specified term of arithmetic sequences
A25	Determine if functions are one to one (using graphs)
A26	Graph simple ellipses

Appendix B: List of Participating Schools

School Name	City	State/ Province	Region
Cardston Elementary School	Cardston	AB	Canada
R. I. Baker School	Coaldale	AB	Canada
Nikiski Middle/High School	Nikiski	AK	W
Romig Middle School	Anchorage	AK	W
Athens High School	Athens	AL	SE
Bankhead Middle School	Cordova	AL	SE
Centre Elementary School	Centre	AL	SE
Cleburne Co Elementary School	Heflin	AL	SE
Clements High School	Athens	AL	SE
Coldwater Elementary School	Anniston	AL	SE
Fleeta School	Opp	AL	SE
Harry N. Mixon Elementary School	Ozark	AL	SE
Headland Elementary School	Headland	AL	SE
Indian Springs Elementary School	Eight Mile	AL	SE
Ruhama School	Fort Payne	AL	SE
Sulligent Consolidated School	Sulligent	AL	SE
Verner Elementary School	Tuscaloosa	AL	SE
Angie Grant Elementary School	Benton	AR	SE
Clarendon Junior/Senior High School	Clarendon	AR	SE
Dardanelle Middle School	Dardanelle	AR	SE
Dover High School	Dover	AR	SE
Harrisburg Elementary School	Harrisburg	AR	SE
Hughes Junior/Senior High School	Hughes	AR	SE
Jessieville Elementary School	Jessieville	AR	SE

School Name	City	State/ Province	Region
Julia Shannon Elementary School	Stuttgart	AR	SE
Mayflower Elementary School	Mayflower	AR	SE
Quitman Junior/Senior High	Quitman	AR	SE
Rison Elementary School	Rison	AR	SE
Rison High School	Rison	AR	SE
Sunnymede Elementary School	Fort Smith	AR	SE
Vilonia Elementary School	Vilonia	AR	SE
White Hall Junior High School	White Hall	AR	SE
Holiday Park Elementary School	Phoenix	AZ	W
Larkspur Elementary School	Phoenix	AZ	W
Richmond Christian School	Richmond	BC	Canada
Arroyo Mocho School	Livermore	CA	W
Browns Valley Elementary School	Vacaville	CA	W
Calvin Christian High School	Escondido	CA	W
Cameron Ranch Elementary School	Carmichael	CA	W
Chana High School	Auburn	CA	W
Credence Continuation High School	Susanville	CA	W
Crestline Elementary School	Barstow	CA	W
Curtner Elementary School	Milpitas	CA	W
El Cajon Valley High School	El Cajon	CA	W
Foster Elementary School	San Diego	CA	W
Futures High School	San Diego	CA	W
Golden Valley High School	Merced	CA	W
Grand View Elementary School	Dinuba	CA	W
Hagginwood Elementary School	Sacramento	CA	W
Kappa High School	Richmond	CA	W
Lake Gregory Elementary School	Lake Crestline	CA	W

School Name	City	State/ Province	Region
Live Oak School	Fallbrook	CA	W
Livermore High School	Livermore	CA	W
Marshall Middle School	San Diego	CA	W
Newark Memorial High School	Newark	CA	W
Northview Intermediate School	Duarte	CA	W
Palla School	Bakersfield	CA	W
Portola Elementary School	Ventura	CA	W
Schallenberger Elementary School	San Jose	CA	W
Shasta Lake Middle School	Shasta Lake	CA	W
Sierra Vista Junior High School	Canyon Country	CA	W
St. Gregory School	San Mateo	CA	W
St. Jerome School	Los Angeles	CA	W
Tompkins Elementary School	Tehachapi	CA	W
Twenty-Fourth Street School	Los Angeles	CA	W
West Covina High School	West Covina	CA	W
Aragon Elementary School	Fountain	CO	W
Bear Creek Elementary School	Lakewood	CO	W
Bergen Valley Elementary School	Evergreen	CO	W
Estes Park Intermediate School	Estes Park	CO	W
Frontier Elementary School	Colorado Springs	CO	W
Goldrick Elementary School	Denver	CO	W
Harrison Elementary School	Canon City	CO	W
Horizon Community Middle School	Aurora	CO	W
Lopez Elementary School	Fort Collins	CO	W
Palmer Lake Elementary School	Palmer Lake	CO	W
Rocky Ford High School	Rocky Ford	CO	W
Summit County Christian School	Frisco	CO	W

School Name	City	State/ Province	Region
Weber Elementary School	Arvada	CO	W
Wildflower Elementary School	Colorado Springs	CO	W
Valley View Elementary School	Denver	CO	W
Harriet Beecher Stowe Elementary School	Enfield	CT	NE
New Britain High School	New Britain	CT	NE
Kirk Middle School	Newark	DE	NE
Paul Laurence Dunbar Elementary School	Laurel	DE	NE
Smyrna High School	Smyrna	DE	NE
Annunciation Catholic Academy	Altamonte Springs	FL	SE
Bonifay Elementary School	Bonifay	FL	SE
Brentwood Elementary School	Pensacola	FL	SE
Cypress Elementary School	Pompano Beach	FL	SE
Emma Love Hardee Elementary School	Fernandina	FL	SE
Ernest Ward School	Walnut Hill	FL	SE
Ethel Koger Beckham Elementary	Miami	FL	SE
Evergreen Elementary School	Ocala	FL	SE
Gulfside Elementary School	Holiday	FL	SE
Horizon Elementary School	Port Orange	FL	SE
Hudson High School	Hudson	FL	SE
J. Colin English Elementary School	North Fort Myers	FL	SE
Marathon Lutheran School	Marathon	FL	SE
New Horizons Learning Center	Panama City	FL	SE
Palmer Catholic Academy	Ponte Vedra Beach	FL	SE
Pelican Elementary School	Cape Coral	FL	SE
Seaside Neighborhood School	Seaside	FL	SE
Seminole Elementary School	Miami	FL	SE
South Dade High School	Homestead	FL	SE

School Name	City	State/ Province	Region
St. Mary School	Fort Walton Beach	FL	SE
Terwilliger Elementary School	Gainesville	FL	SE
White City Elementary School	Fort Pierce	FL	SE
Americus High School	Americus	GA	SE
Appling Middle School	Baxley	GA	SE
East Cobb Middle School	Marietta	GA	SE
Patrick Henry High School	Stockbridge	GA	SE
Pepperell Elementary School	Lindale	GA	SE
Pooler Elementary School	Pooler	GA	SE
R. I. Norton Elementary School	Snellville	GA	SE
Riverside Elementary School	Evans	GA	SE
Southwest Middle School	Savannah	GA	SE
Springfield Central Elementary School	Springfield	GA	SE
Sumter County Middle School	Americus	GA	SE
Toombs County High School	Lyons	GA	SE
Unadilla Elementary School	Unadilla	GA	SE
Enchanted Lake Elementary School	Kailua	HI	W
Holualoa Elementary School	Holualoa	HI	W
Kahakai Elementary School	Kailua Kona	HI	W
Lunalilo Elementary School	Honolulu	HI	W
Dayton Elementary School	Dayton	IA	MW
Denmark Elementary School	Denmark	IA	MW
Dows Community School	Dows	IA	MW
Fort Dodge High School	Fort Dodge	IA	MW
Gehlen Catholic School	Le Mars	IA	MW
Grand Community School	Boxholm	IA	MW
Grinnell Middle School	Grinnell	IA	MW

School Name	City	State/ Province	Region
Hanawalt Elementary School	Des Moines	IA	MW
Harry S. Truman Elementary School	Davenport	IA	MW
Irving Elementary School	Dubuque	IA	MW
Jefferson Elementary School	Fort Madison	IA	MW
Kanesville High School	Council Bluffs	IA	MW
Lincoln School	Fort Madison	IA	MW
Pact Alternative School	Council Bluffs	IA	MW
St. Theresa School	Des Moines	IA	MW
Blackfoot Sixth Grade Center	Blackfoot	ID	W
Canyon Elementary School	Cataldo	ID	W
Council Elementary School	Council	ID	W
Council Junior/Senior High School	Council	ID	W
Genesee Elementary School	Genesse	ID	W
Gooding High School	Gooding	ID	W
Irving Junior High School	Pocatello	ID	W
Kellogg Middle School	Kellogg	ID	W
Lincoln Elementary School	Saint Anthony	ID	W
Lincoln Elementary School	Twin Falls	ID	W
Malad Elementary School	Malad City	ID	W
Meadows Valley School	New Meadows	ID	W
Pinehurst Elementary School	Pinehurst	ID	W
Sunnyside Elementary School	Kellogg	ID	W
Alwood Elementary School	Alpha	IL	MW
Beulah Park School	Zion	IL	MW
Big Hollow School	Ingleside	IL	MW
Chebanse Elementary School	Chebanse	IL	MW
Creal Springs Elementary School	Creal Springs	IL	MW

School Name	City	State/ Province	Region
Crete-Monee Middle School	University Park	IL	MW
Dakota Junior/Senior High School	Dakota	IL	MW
Effingham Junior High School	Effingham	IL	MW
Glen Oak Primary School	Peoria	IL	MW
Grant Elementary School	Melrose Park	IL	MW
Grantfork Elementary School	Highland	IL	MW
Le Vasseur Elementary School	Bourbonnais	IL	MW
Lincoln Middle School	Park Ridge	IL	MW
Marquette Elementary School	Machesney Park	IL	MW
North School	Des Plaines	IL	MW
Pekin Community High School	Pekin	IL	MW
Southeastern Junior High School	Bowen	IL	MW
St. Florian Elementary School	Chicago	IL	MW
St. Hubert School	Hoffman Estate	IL	MW
St. Louise De Marillac School	Lagrange Park	IL	MW
St. Paul The Apostle School	Joliet	IL	MW
St. Philomena School	Peoria	IL	MW
Brookview Elementary School	Indianapolis	IN	MW
Byrkit High School	Mishawaka	IN	MW
Discovery Middle School	Granger	IN	MW
La Porte High School	La Porte	IN	MW
Manchester High School	North Manchester	IN	MW
Nativity Catholic School	Indianapolis	IN	MW
Owen Valley Middle School	Spencer	IN	MW
Rose Hamilton Elementary School	Centerville	IN	MW
St. Mary Cathedral School	La Fayette	IN	MW
Wilson Elementary School	Jeffersonville	IN	MW

School Name	City	State/ Province	Region
Brooks Technology and Arts Magnet School	Wichita	KS	MW
Curtis Middle School	Wichita	KS	MW
Galena High School	Galena	KS	MW
Galena Middle School	Galena	KS	MW
Green Springs Elementary School	Olathe	KS	MW
Holy Family Elementary School	Hays	KS	MW
Hugoton Middle School	Hugoton	KS	MW
Pittsburg High School	Pittsburg	KS	MW
Prairie View Middle School	Lacygne	KS	MW
St. Paul High School	St. Paul	KS	MW
Sunflower Elementary School	Andover	KS	MW
Thayer Elementary/High School	Thayer	KS	MW
Trinity Lutheran School	Atchison	KS	MW
Bald Knob Elementary School	Frankfort	KY	SE
Conner Middle School	Hebron	KY	SE
Corbin Middle School	Corbin	KY	SE
Crab Orchard Elementary School	Crab Orchard	KY	SE
Greenwood High School	Bowling Green	KY	SE
Heath High School	West Paducah	KY	SE
Lincoln Trail Elementary School	Elizabethton	KY	SE
Mary Queen Of Heaven School	Erlanger	KY	SE
Morgan County High School	West Liberty	KY	SE
North Pointe Elementary School	Florence	KY	SE
Oakview Elementary School	Ashland	KY	SE
Rockcastle County High School	Mount Vernon	KY	SE
Sturgis Elementary School	Sturgis	KY	SE
Bissonet Plaza Elementary School	Metairie	LA	SE

School Name	City	State/ Province	Region
Boutte Christian Academy	Boutte	LA	SE
Destrehan High School	Destrehan	LA	SE
Lakeview Junior/Senior High School	Campti	LA	SE
Larose Middle School	Larose	LA	SE
Loyola College Prep School	Shreveport	LA	SE
Marrero Christian Academy	Marrero	LA	SE
Mire Elementary School	Rayne	LA	SE
Northeast High School	Zachary	LA	SE
Peabody Sixth Grade Center	Alexandria	LA	SE
St. Frederick High School	Monroe	LA	SE
St. Gregory Barbarigo School	Houma	LA	SE
St. Ignatius School	Grand Coteau	LA	SE
West Feliciana Middle School	St. Francisville	LA	SE
Helen Mae Sauter School	Gardner	MA	NE
J. B. Devalles Elementary School	New Bedford	MA	NE
Murphy Primary School	Weymouth	MA	NE
Plymouth Community Intermediate School	Plymouth	MA	NE
Prospect School	Gardner	MA	NE
Sheffield Elementary School	Turners Falls	MA	NE
South Hadley High School	South Hadley	MA	NE
Cedarmere Elementary School	Reisterstown	MD	NE
Our Lady Of Mt. Carmel Elementary School	Essex	MD	NE
Our Lady Of Mt. Carmel High School	Essex	MD	NE
Swan Meadow School	Oakland	MD	NE
Washington High School	Princess Anne	MD	NE
Westernport Elementary School	Westernport	MD	NE
Yeshivat Rambam School	Baltimore	MD	NE

School Name	City	State/ Province	Region
East Grand School	Danforth	ME	NE
New Sweden School	New Sweden	ME	NE
Bishop Gallagher High School	Harper Woods	MI	MW
Central Elementary School	Vassar	MI	MW
Garfield-Fraser Elementary School	Pinconning	MI	MW
Gaylord Intermediate School	Gaylord	MI	MW
Lakeview Elementary School	Ludington	MI	MW
Townsend North Elementary School	Vassar	MI	MW
Waylee Elementary School	Portage	MI	MW
Frost Lake School	St. Paul	MN	MW
Heartland Christian Academy	Bemidji	MN	MW
Jefferson Elementary School	Blaine	MN	MW
Kelliher Public School	Kelliher	MN	MW
Kennedy Elementary School	Mankato	MN	MW
Minnesota Learning Center	Brainerd	MN	MW
Princeton High School	Princeton	MN	MW
Randolph Heights Elementary School	St. Paul	MN	MW
St. Elizabeth Seton School	Minneapolis	MN	MW
Upsala School	Upsala	MN	MW
Westwood Elementary School	Blaine	MN	MW
Worthington Middle School	Worthington	MN	MW
Beaufort Elementary School	Beaufort	MO	MW
Bloomfield Elementary School	Bloomfield	MO	MW
Greenfield R-IV Elementary School	Greenfield	MO	MW
Humansville Junior Senior High School	Humansville	MO	MW
Kingston K-14 Middle School	Cadet	MO	MW
Liberty Academy	Liberty	MO	MW

School Name	City	State/ Province	Region
Maysville Elementary School	Maysville	MO	MW
Norborne Elementary School	Norborne	MO	MW
Pettis Co. R 12 Elementary School	Sedalia	MO	MW
Robidoux Middle School	Saint Joseph	MO	MW
Shook Elementary School	Marshfield	MO	MW
St. Clement School	Bowling Green	MO	MW
St. Monica Catholic School	Kansas City	MO	MW
Walker Elementary School	Florissant	MO	MW
Windsor Elementary School	Windsor	MO	MW
Ackerman Junior/Senior High School	Ackerman	MS	SE
Amory High School	Amory	MS	SE
Bel-Aire Elementary School	Gulfport	MS	SE
Centreville Academy	Centreville	MS	SE
East Kemper Elementary	Scooba	MS	SE
Hamilton School	Hamilton	MS	SE
Higgins Middle School	Mccomb	MS	SE
Jackson Academy	Jackson	MS	SE
Jonestown Elementary School	Clarksdale	MS	SE
Lucy Webb Elementary School	Greenville	MS	SE
Magnolia Middle School	Meridan	MS	SE
Northwest Rankin Middle School	Brandon	MS	SE
Thomas Street Elementary School	Tupelo	MS	SE
Velma Jackson Magnet High School	Camden	MS	SE
Butte High School	Butte	MT	W
Fred Moodry Middle School	Anaconda	MT	W
Noxon School	Noxon	MT	W
St. Ignatius Elementary School	St. Ignatius	MT	W

School Name	City	State/ Province	Region
Princess Elizabeth School	Saint John	NB	Canada
Woodlawn Learning Centre	Saint John	NB	Canada
Bath Elementary School	Bath	NC	SE
C. B. Aycock High School	Pikeville	NC	SE
East Clayton Elementary School	Clayton	NC	SE
East Elementary School	Monore	NC	SE
Eastfield Elementary School	Marion	NC	SE
Falkland Elementary School	Greenville	NC	SE
Fremont Stars School	Fremont	NC	SE
Hudson Elementary School	Hudson	NC	SE
Meadow Lane Elementary School	Goldsboro	NC	SE
Montclair Elementary School	Fayetteville	NC	SE
Moss Hill Elementary School	Kinston	NC	SE
Myrtle Grove Middle School	Wilmington	NC	SE
South Lenoir High School	Deep Run	NC	SE
Stanfield Elementary School	Stanfield	NC	SE
Swift Creek Elementary School	Raleigh	NC	SE
Troutman Middle School	Troutman	NC	SE
Tryon Elementary School	Tryon	NC	SE
Valmead Elementary School	Lenoir	NC	SE
Zeb Vance Elementary School	Kittrell	NC	SE
Sunnyside Elementary School	Minot	ND	MW
Johnson-Brock Elementary School	Brock	NE	MW
Lincoln Christian School	Lincoln	NE	MW
Lincoln Elementary School	Norfolk	NE	MW
O'Neill Elementary School	O'Neill	NE	MW
Alton Central School	Alton	NH	NE

School Name	City	State/ Province	Region
Broken Ground School	Concord	NH	NE
Elm Street Elementary School	Laconia	NH	NE
Hillsboro-Deering High School	Hillsboro	NH	NE
Kimball School	Concord	NH	NE
Russell Elementary School	Rumney	NH	NE
Spaulding High School	Rochester	NH	NE
Hightstown High School	Hightstown	NJ	NE
Holy Family Academy	Bayonne	NJ	NE
Mt. Arlington Elementary School	Mt. Arlington	NJ	NE
Pleasantech Academic Charter School	Pleasantville	NJ	NE
Resurrection School	Jersey City	NJ	NE
Richard Butler Middle School	Butler	NJ	NE
Sussex County Charter School	Sparta	NJ	NE
Twin Hills School	Willingboro	NJ	NE
Upper Pittsgrove School	Monroeville	NJ	NE
Washington Avenue School	Chatham	NJ	NE
Chaparrel Elementary School	Deming	NM	W
Holloman Middle School	Alamogordo	NM	W
Mesa Elementary School	Clovis	NM	W
Mountain View Middle School	Alamogordo	NM	W
New Futures Middle/High School	Albuquerque	NM	W
Ojo Amarillo Elementary School	Fruitland	NM	W
Parkview Elementary School	Roswell	NM	W
Picacho Middle School	Las Cruces	NM	W
Tome Elementary School	Los Lunas	NM	W
Tularosa Middle School	Tularosa	NM	W
Jacobson Elementary School	Las Vegas	NV	W

School Name	City	State/ Province	Region
Pau Wa Lu Middle School	Gardnerville	NV	W
Sandy Valley School	Sandy Valley	NV	W
Seeliger Elementary School	Carson City	NV	W
Wells School	Wells	NV	W
William G. Bennett Elementary School	Laughlin	NV	W
Alden Primary School	Alden	NY	NE
Archbishop Walsh High School	Olean	NY	NE
Belfast Central School	Belfast	NY	NE
Bell Elementary School	Kirkwood	NY	NE
East Middle School	West Seneca	NY	NE
F. D. Roosevelt High School	Hyde Park	NY	NE
Freewill Elementary School	Ontario Center	NY	NE
Governor George Clinton Elementary School	Poughkeepsie	NY	NE
Marathon Central School	Marathon	NY	NE
Middleburgh Middle School	Middleburgh	NY	NE
Northstar Christian Academy	Rochester	NY	NE
St. Mark Lutheran School	North Tonawanda	NY	NE
Woodlands High School	Hartsdale	NY	NE
Allen Elementary School	Curtice	OH	MW
Fort Miami Elementary School	Maumee	OH	MW
Hillsboro Middle School	Hillsboro	OH	MW
Longfellow-Emerson Elementary School	Lorain	OH	MW
Montgomery School	Ashland	OH	MW
New Richmond High School	New Richmond	OH	MW
Opportunity School	Wooster	OH	MW
Orange High School	Pepper Pike	OH	MW
Pandora-Gilboa Elementary School	Pandora	OH	MW

School Name	City	State/ Province	Region
Pandora-Gilboa High School	Pandora	OH	MW
Paulding Middle School	Paulding	OH	MW
Robinwood Lane Elementary School	Boardman	OH	MW
Southington Local School	Southington	OH	MW
St. Clairsville Elementary School	St. Clairsville	OH	MW
St. Francis De Sales School	Newark	OH	MW
St. Francis Xavier School	Willard	OH	MW
St. Jude School	Elyria	OH	MW
St. Mary School	Delaware	OH	MW
Stivers School For The Arts	Dayton	OH	MW
Swiss Hills Career Center	Woodsfield	OH	MW
West Hill Elementary School	Rittman	OH	MW
Academy Elementary School	Guymon	OK	W
Carl Albert Junior High School	Oklahoma City	OK	W
Clark Elementary School	Tulsa	OK	W
Coronado Heights Elementary School	Oklahoma City	OK	W
Fargo School	Fargo	OK	W
Fox K-12 School	Fox	OK	W
Harrah Middle School	Harrah	OK	W
Hugh Bish Elementary School	Lawton	OK	W
Lookeba-Sickles Elementary School	Lookeba	OK	W
Madison Middle School	Bartlesville	OK	W
Merritt Elementary School	Elk City	OK	W
Northeast Elementary School	Guymon	OK	W
Pryor Alternative School	Pryor	OK	W
Robert E. Lee Elementary School	Duncan	OK	W
Rockwood Elementary School	Oklahoma City	OK	W

School Name	City	State/ Province	Region
Southside Elementary School	Broken Arrow	OK	W
Tyrone School	Tyrone	OK	W
Watongo Middle School	Watonga	OK	W
Windsor Hills Elementary School	Oklahoma City	OK	W
Sir Mackenzie Bowell School	Belleville	ON	Canada
Jewett Elementary School	Central Point	OR	W
Laurelhurst Elementary School	Portland	OR	W
Manzanita Elementary School	Grants Pass	OR	W
Patrick Elementary School	Gold Hill	OR	W
Westridge Middle School	Westfir	OR	W
Benton Junior/Senior High School	Benton	PA	NE
Cambria County Christian School	Johnstown	PA	NE
Dunmore Elementary School	Dunmore	PA	NE
Gettysburg Area Middle School	Gettysburg	PA	NE
Immaculate Heart School	Girardville	PA	NE
L. R. Appleman Elementary School	Benton	PA	NE
Mayfair Elementary School	Philadelphia	PA	NE
Mowrey Elementary School	Waynesboro	PA	NE
Paxtonia Elementary School	Harrisburg	PA	NE
Springfield Elementary School	East Springfield	PA	NE
St. Edward School	Herminie	PA	NE
St. Germaine School	Bethel Park	PA	NE
Town and Country Day School	Harrisburg	PA	NE
Burrillville High School	Harrisville	RI	NE
Brittons Neck Elementary School	Britton's Neck	SC	SE
Buford High School	Lancaster	SC	SE
C. E. Murray Junior/Senior High School	Greeleyville	SC	SE

School Name	City	State/ Province	Region
Camden Middle School	Camden	SC	SE
Clinton Elementary School	Clinton	SC	SE
Concord Elementary School	Anderson	SC	SE
Fair Oak Elementary School	Westminster	SC	SE
Goodwin Elementary School	North Charleston	SC	SE
Lambs Elementary School	North Charleston	SC	SE
Mabry Junior High School	Inman	SC	SE
Marshall Primary School	Belton	SC	SE
Mitchell Road Elementary School	Greenville	SC	SE
Orangeburg-Wilkinson High School	Orangeburg	SC	SE
Pee Dee Elementary School	Conway	SC	SE
Lincoln Elementary School	Watertown	SD	MW
St. Mary High School	Dell Rapids	SD	MW
Invermay School	Invermay	SK	Canada
Barger Academy of Fine Arts	Chattanooga	TN	SE
Blue Springs Elementary School	Cleveland	TN	SE
Cason Lane Academy	Murfreesboro	TN	SE
Denver Elementary School	Memphis	TN	SE
East Ridge Middle School	Chattanooga	TN	SE
Frayser High School	Memphis	TN	SE
Freedom Intermediate School	Franklin	TN	SE
Griffith Elementary School	Dunlap	TN	SE
Ingram Sowell Elementary School	Lawrenceburg	TN	SE
Loudon Elementary School	Loudon	TN	SE
Madisonville Primary School	Madisonville	TN	SE
McDowell Elementary School	Columbia	TN	SE
McConnell Elementary School	Hixson	TN	SE

School Name	City	State/ Province	Region
Mt. Juliet Elementary School	Mt. Juliet	TN	SE
Richland Elementary School	Lynnville	TN	SE
Shelby Oaks Elementary School	Memphis	TN	SE
Trenton Rosenwald Middle School	Trenton	TN	SE
Valley Pike Elementary School	Bristol	TN	SE
W. A. Wright Elementary School	Mt. Juliet	TN	SE
Washburn School	Washburn	TN	SE
West Lafollette Elementary School	Lafollette	TN	SE
Whitwell Middle School	Whitwell	TN	SE
Bennett Intermediate School	Wolfforth	TX	W
Blooming Grove Elementary School	Blooming Grove	TX	W
Bondy Intermediate School	Pasadena	TX	W
Bowie Elementary School	Sulphur Springs	TX	W
Brown Elementary School	Whitehouse	TX	W
Burnet Elementary School	Odessa	TX	W
Burton Hill Elementary School	Fort Worth	TX	W
Canyon Intermediate School	New Braunfels	TX	W
Clifton Middle School	Houston	TX	W
Corrigan-Camden Elementary School	Corrigan	TX	W
Edcouch-Elsa Junior High School	Edcouch	TX	W
Eisenhower Middle School	San Antonio	TX	W
Franklin Elementary School	Port Arthur	TX	W
George West Elementary School	George West	TX	W
Goldthwaite Elementary School	Goldthwaite	TX	W
Hays Elementary School	Odessa	TX	W
Hillcrest Elementary School	Nederland	TX	W
Holloway Middle School	Whitehouse	TX	W

School Name	City	State/ Province	Region
Jefferson Elementary School	Sherman	TX	W
Jones Elementary School	Corpus Christi	TX	W
Joshua Accelerated Learning Center	Joshua	TX	W
Little Cypress-Mauriceville School	Orange	TX	W
Martin Special Emphasis Middle School	Corpus Christi	TX	W
Maurine Cain Middle School	Rockwall	TX	W
Midland Freshman High School	Midland	TX	W
New Summerfield Independent School District	New Summerfield	TX	W
Owens Elementary School	Tyler	TX	W
Patterson Literature Magnet School	Houston	TX	W
Paul R. Haas Middle School	Corpus Christi	TX	W
Popham Elementary School	Del Valle	TX	W
Ridgewood Elementary School	Port Neches	TX	W
St. Mary's Central Catholic School	Odessa	TX	W
Tom Bean Elementary School	Tom Bean	TX	W
Travis Elementary School	Memphis	TX	W
Trimble Tech High School	Fort Worth	TX	W
Walnut Bend Elementary School	Houston	TX	W
Wheatley High School	Houston	TX	W
Wortham Elementary School	Wortham	TX	W
Xenia Voigt Elementary School	Round Rock	TX	W
Cook Elementary School	Syracuse	UT	W
Discovery Elementary School	Vernal	UT	W
Farmington Bay Youth Center	Farmington	UT	W
Grouse Creek School	Grouse Creek	UT	W
Parkview Elementary School	Salt Lake City	UT	W
San Rafael Junior High School	Ferron	UT	W

School Name	City	State/ Province	Region
Silver Hills Elementary School	Kearns	UT	W
St. Joseph Catholic High School	Ogden	UT	W
Bethel Elementary School	Gloucester	VA	SE
Bland School	Bland	VA	SE
Boones Mill Elementary School	Boones Mill	VA	SE
Dearington Elementary School	Lynchburg	VA	SE
Fairfield Court Elementary School	Richmond	VA	SE
Fishburne Military School	Waynesboro	VA	SE
Greendale Elementary School	Abingdon	VA	SE
J. W. Adams Elementary	Pound	VA	SE
Malibu Elementary School	Virginia Beach	VA	SE
Marshall Middle School	Marshall	VA	SE
Orange County High School	Orange	VA	SE
Reservoir Middle School	Newport News	VA	SE
Stone Spring Elementary School	Harrisburg	VA	SE
Warrenton Middle School	Warrenton	VA	SE
William Byrd Middle School	Vinton	VA	SE
Fleming Elementary School	Essex Junction	VT	NE
Washington Village School	Washington	VT	NE
Cedarcrest School	Marysville	WA	W
Clover Valley Elem School	Oak Harbor	WA	W
East Ridge Elementary School	Woodinville	WA	W
Gates High School	Tacoma	WA	W
Kapowsin Elementary School	Graham	WA	W
Lincoln Hill High School	Stanwood	WA	W
Oakview Elementary School	Centralia	WA	W
Pace Alternative School	Wapato	WA	W

School Name	City	State/ Province	Region
Pratt Elementary School	Spokane	WA	W
R. E. Bennett Elementary School	Chehalis	WA	W
Stanley Magnet School	Tacoma	WA	W
Vale Elementary School	Cashmere	WA	W
West View School	Burlington	WA	W
Audubon Middle School	Milwaukee	WI	MW
Aurora Weier Education Center	Milwaukee	WI	MW
Evergreen Elementary School	Mosinee	WI	MW
Frank Allis Elementary School	Madison	WI	MW
Johnson Creek Elementary School	Johnson Creek	WI	MW
Lance Middle School	Kenosha	WI	MW
Mead Elementary School	Wisconsin Rapids	WI	MW
Mosinee Middle School	Mosinee	WI	MW
North Division High School	Milwaukee	WI	MW
Oostburg Christian School	Oostburg	WI	MW
Pittsville Elementary School	Pittsville	WI	MW
Prairie River Middle School	Merrill	WI	MW
Roberts Elementary School	Fond Du Lac	WI	MW
Seton Catholic Middle School	Menasha	WI	MW
St. Peter Lutheran School	Chilton	WI	MW
St. Martin Lutheran School	Clintonville	WI	MW
Steuben Middle School	Milwaukee	WI	MW
Stratford Elementary School	Stratford	WI	MW
Vesper Elementary School	Vesper	WI	MW
Wakanda Elementary School	Menomonie	WI	MW
Westside Academy II	Milwaukee	WI	MW
Wileman Elementary School	Delavan	WI	MW

School Name	City	State/ Province	Region
Wisconsin Career Academy	Milwaukee	WI	MW
Work and Learning Center	Madison	WI	MW
Yahara Elementary School	Deforest	WI	MW
Central Elementary School	Beckley	WV	SE
Franklin Primary School	Wellsburg	WV	SE
Glenwood Elementary School	Charleston	WV	SE
Marion County Alternative School	Fairmont	WV	SE
Mount Hope High School	Mount Hope	WV	SE
Preston High School	Kingwood	WV	SE
South Preston Middle School	Tunnelton	WV	SE
St. John School	Wellsburg	WV	SE
Valley High School	Smithers	WV	SE
Kemmerer Elementary School	Diamondville	WY	W
Lovell Elementary School	Lovell	WY	W

Index

Index

D

E

N

O

P

Q

R

S

T

U

V

W

Renaissance Learning
P.O. Box 8036, Wisconsin Rapids, WI 54495-8036
(866) 492-6284
Fax: (715) 424-4242
Email: answers@renlearn.com
Web: www.renlearn.com

300219