Perfect for RTI

Getting the Most Out of STAR Math™

Using Data to Inform Instruction and Intervention

Renaissance Learning™

Introduction

STAR Math is a computer-adaptive assessment designed to give you accurate, reliable, and valid data quickly so that you can make good decisions about instruction and intervention. STAR Math, operating on the Renaissance Place Real Time platform, is part of a breakthrough comprehensive assessment system for data-driven schools, which also includes STAR Reading and STAR Early Literacy.

The purpose of this book is to help teachers and administrators get the most out of STAR Math. We begin by explaining the test's design and the types of data generated by the test. We also briefly describe fundamental psychometric attributes of STAR Math, such as validity and reliability. In later chapters, we explain how to best use the test for screening and progress monitoring, and we answer general frequently asked questions. In the appendix, we provide instructions for common software tasks. To make the book useful to a wide audience of educators, we minimize technical terms while explaining the concepts that are important to know. (The STAR Math software contains a technical manual for anyone who wants to examine the psychometric data more closely. You can also find the technical manual in the Resource Center on our Web site's Training Center at www.renlearn.com/profdevel.)

We believe STAR Math is the perfect tool for data-driven schools. It is practical and sound, and it provides a wealth of information about your students' math ability. We hope the information you find here will help and inspire you. It is, however, only an introduction. To learn about more professional-development opportunities, including consultation on your own student data, visit our Web site's Training Center at www.renlearn.com/profdevel.

1

STAR Math Basics

The only way to know whether learning is taking place is to measure it. Once you do that you can do a host of other things. You can provide students with appropriate materials. You can identify students who need help. You can analyze problems with individuals, grades, or schools; set learning goals; and make plans for meeting those goals. And you can determine whether the instruction and intervention you provide are effective.

STAR Math is uniquely capable of facilitating all these tasks. Thanks to computer-adaptive technology, students typically complete the test in less than 15 minutes, and teachers and administrators receive the results immediately. Moreover, STAR Math is accurate, reliable, and valid. In fact, it is highly rated for both screening and progress monitoring by the National Center on Response to Intervention.

In this chapter, we tell you for whom STAR Math is designed, how it works, the types of data it generates, and how we know it is a good assessment. In later chapters, we explain how you can use STAR Math throughout the school year to make thoughtful decisions that will accelerate learning for all of your students.

For Whom Is STAR Math Designed?

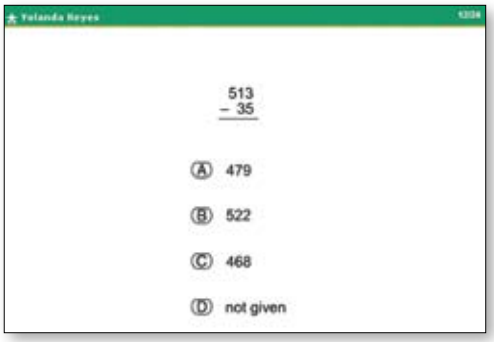
STAR Math is designed for students in grades 1 through 12, but can also be used with kindergarten students. It identifies students' instructional math levels and compares their math achievement to that of students across the nation. The test provides norm-referenced and criterion-referenced scores for students in grades 1 through 12; kindergarten students only receive criterion-referenced scores.

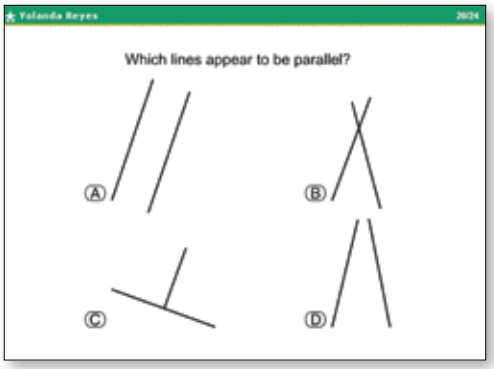
How do you know if a student is ready to take STAR Math? In general, the student should have beginning reading skills and some math skills. Observe the student working through the practice questions that appear at the beginning of STAR Math. If the student can answer these questions unassisted, he is ready to complete the test. If the student has a lot of trouble getting through the practice, the student probably does not have the basic skills necessary to be measured by STAR Math.

Test Frequency

Most schools administer the test at least twice—in fall and spring—to get baseline data for each student and to measure growth over the school year. Many schools test more frequently. They use STAR Math for screening purposes in fall, winter, and spring, and they monitor the progress of students in intervention programs more frequently—perhaps monthly, biweekly, or even weekly.

Test Content





STAR Math's content is organized into eight mathematical strands: numeration concepts, computation processes, word problems, estimation, data analysis and statistics, geometry, measurement, and algebra. The eight strands cover 214 objectives and were identified after consulting multiple sources, including textbook series, state curriculum guides, Principles and Standards for School Mathematics of the National Council of Teachers of Mathematics (NCTM), and Trends in International Mathematics and Science Study (TIMSS). As a result, the content reflects the objectives commonly taught in the math curriculum of contemporary schools (primarily in the United States). The technical manual includes more detailed information about each of the strands.

STAR Math tests consist of 24 questions, not including practice questions and a few items that are in the calibration process, and they are divided into two main parts. The first part, or the "heart" of the test, assesses numeration concepts (items 1–8) and computation processes (items 9–16). The "heart" is the bulk of the test since it covers the two mathematical strands that are fundamental to all others and includes the content about which teachers often desire the most information. The second part, or the "applications," covers all of the remaining strands: word problems, estimation, data analysis and statistics, geometry, measurement, and algebra. The specific makeup of the "applications" depends, in part, on the student's grade level. For example, a fifth-grade student will receive at least one estimation item, but a first-grade student won't receive any. The technical manual details the number of items from each strand that students, by grade level, could possibly see on the test.

STAR Math test items are presented in a multiple-choice format with four answer choices. Students will only see a "not given" response option for items in the computation processes strand. This option was included in order to minimize estimation as a response strategy and to encourage students to actually work the problems to completion.

How STAR Math Works

Students take STAR Math tests at individual computers. The software delivers multiple-choice items one by one, and the student selects answers by using four letter keys (A, B, C, D) and the Enter key (or return key for Macintosh computers). Students follow a protocol: They can use blank scratch paper and a pencil while testing, but not calculators or reference materials. After the test is completed, the software calculates a score, and teachers and administrators view and analyze reports that show results for an individual, class, grade, or school.

STAR Math can provide accurate data in a short amount of time because it combines cutting-edge computer-adaptive technology with a specialized psychometric test design. The best way to understand how this works is to walk through the test-taking experience.

Students start the test. You begin by explaining the test to your students using the pretest instructions that are printed from the software. These instructions explain what the test looks like, how to answer questions, and what happens if a student doesn't answer a question in the time allowed. Each student then takes the test at a computer by logging in with a unique user name and password that you obtain by printing the Student Information Report. (See the appendix for instructions.) The software presents practice questions first and, if the student does fine with those, begins delivering actual test questions.

The first actual test question will be at a level slightly below the student's grade level. Or, if the student has taken a STAR Math test in the previous 180 days, the software uses the results from the last test to determine the starting difficulty level for the next one. You also can adjust a student's starting level in the software by entering a math instructional level, perhaps based on a score from a different test or your professional judgment. If so, the software will start the test with a question at a level slightly below the one you entered. (Instructions for adjusting a student's starting level are in the appendix.)

Testing Conditions

For STAR Math test results to be valid, students must test under certain conditions. Students should have blank scratch paper and pencils to work problems, but not calculators or other reference materials. Also, students should not treat STAR Math like a game, even though they take it at a computer. For this reason, the best setting for testing is often a quiet computer lab in which all students can be tested at once and the testing environment can be carefully controlled. In addition, the pretest instructions found in the software must be read aloud prior to each student's first STAR Math test. In fact, to periodically remind students of testing procedures, it's a good idea to use the instructions during each screening period. For any additional testing beyond the screenings, use your judgment as to whether a student needs to hear the instructions again. See the appendix for information on locating and printing the pretest instructions.

The software adjusts the difficulty of every item. During the actual test, the software uses student responses to determine the level of difficulty of next items. If the student answers an item correctly, the software bumps up the difficulty level of the next item. If the student answers incorrectly, the software lowers the difficulty level of the next item. The same thing happens with the next item and the next. By continually adjusting the difficulty of an item to what the student has shown she can or cannot do, the software zeroes in on an accurate assessment of ability.

We use a similar procedure in our everyday lives. As an example, let's suppose you are new to weight lifting. Perhaps you read in a fitness book that the average person of your age and gender can comfortably lift 10-pound dumbbells overhead. When you try it, those 10 pounds are easy! So you attempt 30 pounds. But, uh-oh, that's too hard. Next you lift 20 pounds—still too hard. After a little more trial and error, you conclude that 14 pounds is just right. Thus, your current ability for lifting dumbbells overhead is 14 pounds.

STAR Math uses the same kind of procedure. The software stores a huge number of items calibrated for difficulty and "adapts" the test to each individual.

Students are given a specific amount of time to answer each question. Time limits keep the test moving and maintain test security, and were determined based on data we obtained when validating the test. Students have up to three minutes to answer each item. You have the option of extending time limits for individual students who you believe need more time to answer each question—English language learners, for example, or students with certain disabilities. Those students will then have twice as long to answer each question. Keep in mind that norms as well as other technical data, such as reliability and validity, are based on administering the test using the standard time limits. Therefore, if you do extend time limits for students, be sure to interpret their norm-referenced scores with caution.

Regardless of the time-limit setting, students receive a warning when 15 seconds remain for answering an item. Items that time out are counted as incorrect unless the student has already selected the correct answer.

The test stops after the student answers 24 questions. A major challenge when testing students is gathering enough evidence to draw reliable conclusions about their ability. This is especially problematic when designing conventional tests. Because every student takes the same test form, a conventional test must contain a large number of items in order to evaluate a wide spread of abilities.

Each STAR Math test, on the other hand, is individualized and unique. Because it immediately adjusts to each student's math ability, it delivers an accurate and reliable score after only 24 questions (not including the practice questions and a few items that are in the calibration process).

The software calculates a score. To measure someone's ability to do a task, you must know how difficult the task is to do. For example, think again about how you determine your weight-lifting ability. You need items—the dumbbells—and a way to express their relative weight, which is called a scale. In this case, the scale is "pounds." You identify the relative weight of the dumbbells by marking them with a number along that scale: 3 pounds, 5 pounds, 7 pounds, 10 pounds, and so on.

As we developed STAR Math, we approached test items in the same way. We administered the items to large, nationally representative samples of students, collected the responses, and performed a statistical analysis to determine the difficulty of each item. Using a scale, we marked each item with a difficulty level: 1.67, 1.68, and so on. This process is called item calibration. Currently, we calibrate continuously by including a few additional items on each STAR test, which is why the tests your students take may have 25 or 26 items instead of 24. (Answers for these extra items do not affect a student's score.)

The method of statistical analysis we use is based on Item Response Theory (specifically the Rasch model). This type of analysis relates the probability of a student correctly answering an item to the student's ability and the difficulty of the item. We can get a sense of how this works by returning to our weight-lifting analogy. Let's suppose we asked a large, nationally representative sample of adults to lift dumbbells of varying weights. After analyzing the data, we could say, for example, that the typical 50-year-old female has a 50-50 chance of lifting 10 pounds overhead, a 70-year-old female has a 50-50 chance of lifting 5 pounds overhead, and so on. If you're a 70-year-old female and you can lift 20 pounds overhead, we now have a good idea of your ability! We also know that if you can

lift 20 pounds, you can lift 15 or 10 or 5. In other words, we can predict what you can do without even asking you to do it.

STAR Math can provide the same kind of information. We know a student's grade level, and we know how difficult each item is for each student, given the time of school year and perhaps an initial estimate of math ability. Therefore we can look at a student's pattern of right and wrong answers on a STAR test and provide a statistically sound estimate of the student's ability. We also know the probability of a student answering any item correctly without presenting that item to the student.

The software reports various types of scores. The most important score that the STAR Math software reports is called the **scaled score (SS)**. This score is similar to pounds in our weight-lifting example. It's a fundamental measure that you can use to see growth over time. Just as your weight-lifting ability might increase from 20 pounds to 25 pounds, a student's math ability might grow from 200 to 300. The drawback to scaled scores is that they don't tell you whether a score is good, bad, or middle of the road. As an educator, you need more information in order to know what test results mean and how to respond to them.

For this reason, we calculate additional scores, all of which are derived from scaled scores. These include both criterion-referenced and norm-referenced scores. We'll explain a few of them here and go into more detail in succeeding chapters.

A criterion-referenced score tells you how a student is doing relative to a set standard. In the case of STAR Math, we include **objective clusters** on the Diagnostic Report to show each student's level of proficiency within numeration and computation objectives. (See the example on p. 18.) If a cluster, such as "hundreds," is shaded, it means the student "mastered" it. In other words, the student's ability estimate suggests that she could answer 70 percent or more correct on that cluster. Further, if she mastered "hundreds," then we know she must have mastered "tens" and "ones." That's because content in the numeration concepts and computation processes strands is organized in a hierarchical structure, reflecting the fact that students' understanding of harder mathematical concepts depends upon their understanding of the more basic ones. The solid black line points to the objective cluster that the student is currently developing, or the lowest objective that she has not yet mastered. See the appendix for a list of the objectives included the numeration concepts and computation processes strands.

While criterion-referenced scores are based on what students should be able to do according to a set standard, norm-referenced scores relate to what students across the nation *can actually do*. **Percentile rank (PR)** is one of these scores. A student with a percentile rank of 85, for example, performs as well or better than 85 percent of students nationwide of the same grade at the same time of year.

Another norm-referenced score—and one that is often misunderstood—is **grade-equivalent (GE)**. This, too, compares a student's test performance to that of students nationally. For example, a GE of 4.2 means a student scored as well on STAR Math as did the typical student who is in the second month of fourth grade.

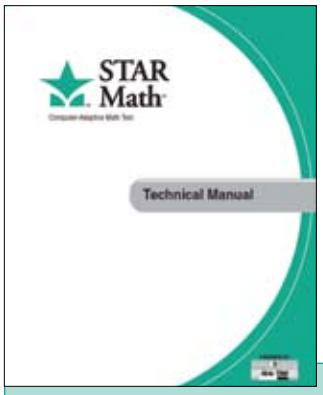
Because different kinds of scores tell us different things, they prompt different actions. Let's look at two more students as examples. We'll say that they were

tested in September and December. The first student, Jennifer Brown, had a scaled score of 704 in September. By December, it increased to 740—she made gains. This is called absolute growth. When we look at her percentile rank, we see that it has also increased—from the 65th percentile to the 74th percentile. This tells us she has made gains relative to her peers, which is called relative growth.

Now let's look at the second student, John Smith. His scaled score has also increased— from 609 to 625—showing absolute growth. But his PR is the same in December as it was in September: the 25th percentile. There's been no relative growth. Thus we know that while John is learning, he's barely maintaining his standing. His growth rate needs to accelerate, perhaps through an intervention program, if he is to make more significant gains.

STAR Math provides many reports that use these and other scores to help you analyze student needs, make good decisions, and monitor progress. We'll give details and examples throughout the rest of this book of the ones that are most commonly used. A list of all the reports available and what they include is in the appendix.

How We Know STAR Math Is a Good Assessment



For additional information, the STAR Math technical manual is available in the software or in the Resource Center on our Web site's Training Center at www.renlearn.com/profdevel.

For a test to be good it must be reliable. A reliable test is like a reliable car. Just as a reliable car starts up every time you turn the key, a reliable test gives consistent results from one administration to another.

With conventional assessments, the key to reliability is length. As we noted earlier, conventional tests must be long in order to provide enough items to adequately test students with a wide range of abilities. Because STAR Math individualizes each test through computer-adaptive technology, it shows high levels of reliability with far fewer items.

Psychometricians evaluate reliability in a number of ways. One way is to administer different, but parallel, tests twice to the same students and see if the scores are consistent. This is referred to as alternate forms reliability. According to the National Center on Response to Intervention (NCRTI), a reliability level of .60 and higher is good; .80 is very good. We have collected and analyzed three types of reliability data, including alternate forms reliability. In all types of analysis, the reliability level of STAR Math exceeds .90.

Besides being reliable, a test must be valid. Validity means that the test actually tests what it is meant to test. As with reliability, there are many ways to measure this. We already looked at STAR Math's "content validity," that is, the relevance of the questions, strands, and objectives sampled by the test. Another way to evaluate validity is to examine the degree to which one assessment correlates with other commonly accepted assessments. To check this, we asked schools to submit students' STAR Math results along with their scores on other assessments, such as the California Achievement Test, Comprehensive Test of Basic Skills, Iowa Test of Basic Skills, Metropolitan Achievement Test, Stanford Achievement Test, and several statewide tests. Our analysis showed a correlation with these tests that exceeded the guideline provided by NCRTI. The technical manual provides details.

Summary

STAR MATH BASICS

- STAR Math is designed for students in grades 1–12 who have beginning reading skills and some math skills. Students in kindergarten may also take the test, but will not receive norm-referenced scores.
- The test is typically administered in fall, winter, and spring for screening purposes and can be administered more frequently for progress monitoring purposes—monthly, biweekly, or even weekly.
- The software adjusts the difficulty of each item to a student's performance. Students typically finish the test in less than 15 minutes.
- The software calculates various scores, including scaled scores, levels of proficiency with numeration and computation objectives, percentile ranks, and grade equivalents, which are used for different purposes.
- STAR Math exceeds standards for reliability and validity.