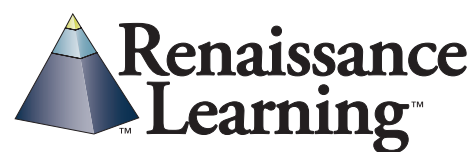


Perfect For
RTI

Getting the Most out of **STAR Reading**[™]

USING DATA TO INFORM INSTRUCTION AND INTERVENTION



Introduction

STAR Reading 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 Reading, operating on the Renaissance Place Real Time platform, is part of a breakthrough comprehensive assessment system for data-driven schools, which also includes STAR Math and STAR Early Literacy.

The purpose of this book is to help teachers and administrators get the most out of STAR Reading. We begin with an explanation of the test's design, the kind of data it generates, and its fundamental psychometric attributes. In later chapters, we explain how to best use the test for screening and progress monitoring. We also answer frequently asked questions and 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. (STAR Reading software contains a technical manual for anyone who wants to examine the psychometric data more closely.)

We believe STAR Reading is the perfect tool for data-driven schools. It is practical and sound, and it provides a wealth of information about your students' reading 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.

1

STAR Reading 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 is effective.

STAR Reading is uniquely capable of facilitating all these tasks. Thanks to computer-adaptive technology, students complete the test in about ten minutes, and teachers and administrators receive the results immediately. Moreover, STAR Reading is accurate, reliable, and valid. In fact, it received the highest rating of all screening assessments from the National Center on Response to Intervention, and is among the highest rated progress-monitoring assessments.

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

For Whom Is STAR Reading Designed?

STAR Reading is designed for students who can read independently. It measures students' reading comprehension and compares their reading achievement to that of students across the nation. The test provides norm-referenced scores for students in grades 1 through 12; kindergarten students who have begun to read may take the test, but norm-referenced scores are not reported.

How do you know if a student is ready to take STAR Reading? A general rule of thumb is that he or she must have a sight vocabulary of about 100 words. You can determine this in a couple of ways.

1. Estimate the student's sight vocabulary based on what you see in the classroom. Is the student able to read early-reader books independently? When called on, can the student read simple text out loud?
2. Give the student the STAR Early Literacy assessment. If the student is identified as a "Probable Reader," he or she can likely take STAR Reading and receive a score.
3. Observe the student working through the practice questions that appear at the beginning of STAR Reading. If the student can answer these questions unassisted, he or she is ready to complete the test.

Testing Emergent Readers

If a student does not have a sufficient vocabulary to take STAR Reading, administer STAR Early Literacy instead. This is a computer-adaptive test for emergent readers and provides proficiency data on 41 early-literacy skills and concepts. In many classrooms, teachers find that some students may take STAR Reading while others must be assessed with STAR Early Literacy.

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 Reading for screening purposes in fall, winter, and spring, and they monitor the progress of students in intervention programs with weekly, biweekly, or monthly testing.

How STAR Reading Works

Students take STAR Reading at individual computers. The software delivers multiple-choice items one by one, and the student selects answers. 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 Reading 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. He or she logs in with a unique username and password that you obtain by printing the Student Information Report. (See the appendix for instructions.) Then the software looks for an estimate of the student's reading ability. If the student took a STAR test within the previous 180 days, the computer refers to that score. Otherwise, the software uses other information that has already been entered. This might be the student's grade placement or an estimate of reading ability based on a score from a different test or your professional judgment. The software presents practice questions first and, if the student does fine with those, begins delivering actual test questions.

Testing Conditions

Because students take STAR Reading at a computer, they sometimes treat it like a game. But for test results to be valid, the test must be taken seriously. For this reason, the best setting is a quiet computer lab in which all students can be tested at once and the testing environment can be carefully controlled. In addition, you must read the standard pretest instructions, which you can print from the software. See the appendix for information on locating them.

The software adjusts the difficulty of every item. After the practice session, the software delivers a “real” test item based on the student's estimated ability level. If the student answers the 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 Reading uses the same kind of procedure. The software stores a huge number of items and “adapts” the test to each individual.

Students are given a specific amount of time to answer each question. Based on data we obtained when validating the test, we have set time limits for test items. Students in grades K-2 have up to 60 seconds to answer each item. Students in grades 3-12 are allowed 45 or 60 seconds, depending on the item type. You have the option of extending time limits for individual students who you believe need more time to read and answer each question—English language learners, for example, or some students with disabilities. Those students will then have three times longer to answer each question. If you do extend the limits, be aware that norms as well as other technical data, such as reliability and validity, are based on administering the test using the standard time limits.

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 25 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 Reading test, on the other hand, is individualized and unique. Because it immediately adjusts to each student’s reading ability, it delivers an accurate and reliable score after only 25 questions (not including the practice questions and a few items that are in the calibration process). In general, the test as a whole takes about 10 to 15 minutes.

The software calculates a score. To report 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 Reading, 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 26 or 27 items instead of 25. (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 Reading can provide the same kind of information. We know a student's grade level, and we know how difficult each item in our item bank is for each student in that grade level. 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 STAR Reading software reports is called the **scaled score**. 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 reading 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 Reading, we look at commonly known vocabulary lists that indicate what words students should know per grade. By comparing a student's test performance to these lists we estimate the student's **Instructional Reading Level (IRL)**. This score represents the highest grade level at which the student can comprehend 80 percent of the text, and is the level at which the student can be most effectively taught. For example, a student with an IRL of 4.5 can comprehend 80 percent of the text that students at the fifth month of fourth grade are expected to comprehend, according to commonly accepted standards.

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 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 is reading at a level comparable to students nationwide who are in the second month of fourth grade.

Often a student's GE and IRL are different. That's because the level at which students perform is not always the same as the level at which we want them to perform as expressed in set standards. Let's suppose a third-grade student has

an Instructional Reading Level of 3.5, indicating he can handle text written at a 3.5 level with 80 percent comprehension. His GE might be somewhat higher—let's say, 4.0. That would tell us that having the ability to read text at a 3.5 level with instructional assistance is actually about average for fourth-graders at the beginning of the school year.

We see these kinds of differences in everyday life, too. For example, the President's Health Commission might set a standard that says 10-year-old children should be able to do thirty jumping jacks without stopping. However, we are a nation of couch potatoes! Someone might do a study and find that the average 10-year-old child can only do fifteen jumping jacks in a row. Thus a child capable of doing fifteen jumping jacks would look pretty weak when compared to the standard, but about average compared to the norm.

Estimated oral reading fluency (ORF) is an estimate of a student's ability to read words quickly and accurately, which in turn leads to efficient comprehension. It is reported as the estimated number of words in grade-level text that the student can read correctly within a one-minute time span. For example, a score of 60 for a second-grade student means the student is expected to correctly read 60 words within one minute on a passage with a readability level between 2.0 and 2.5. Estimated ORF scores are based on the results of a large-scale research study that investigated the links between STAR Reading performance and assessments of oral reading fluency. They are only reported for students in grades 1 – 4. To see a document that identifies cut points and benchmarks for oral reading fluency, scroll to the STAR Reading tab within the software and click Resources.

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 in September of 782. By December, it increased to 946—she definitely made gains. This is called absolute growth. When we look at her percentile rank, we see that it has also increased—from the 57th percentile to the 64th 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 573 to 601—showing absolute growth. But his PR is the same in December as it was in September: the 23rd 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 Reading 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 STAR Reading Measures Comprehension

STAR Reading test items are presented in two formats. In one format, students read a single sentence with a blank to indicate a missing word. The student chooses the word that will complete the sentence from a list of three or four

words. The other format consists of a multi-sentence passage taken from authentic children's literature or nonfiction text. One sentence in the passage contains a blank to indicate a missing word. Here, too, the student must read the passage and choose the word that will complete it from a list of three or four words.

Students in grades K through 2 are presented with 25 single-sentence items. Students in grades 3 through 12 are presented with 20 single-sentence items and five passages. Questions developed at a kindergarten and first-grade reading level show three answer choices. Questions developed at a second-grade level and higher show four answer choices.

While the format of STAR Reading items sometimes leads educators to believe that it only assesses vocabulary knowledge, the test actually draws on a much more complex set of reading skills. Each item is carefully constructed so that the correct answer fits both the semantics and the syntax of the sentence. The incorrect options either fit the syntax of the sentence or relate to the meaning of something in the sentence, but they do not do both. Thus the test-taker must not only apply vocabulary knowledge but must also utilize background knowledge and semantic and syntactical skills. Only if the student uses all these cognitive skills can he derive meaning from the text, which, experts agree, is the essence of reading comprehension.

Nicole Gattari 9 / 25

A week after finishing her first draft, the writer sat down to _____ her work.

1. abolish
2. revise
3. transplant
4. overlap

Taylor Andrews 21 / 25

Long distance runners try to run at a pace which will tire the other runners. To do this, they need plenty of stamina, built up by years of training. They must also know the exact moment to use their speed to break away from and thus upset or confuse the other runners. An _____ with less natural speed than rivals may speed up the pace at any time. Most often, they will do this in the middle of the race.

1. authority
2. inventor
3. errand
4. athlete

How We Know STAR Reading Is a Good Assessment

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.

In the assessment field, 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 Reading 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 the test to the same students within about a week's time and see if the scores are consistent. This is referred to as test-retest 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 four types of reliability data, including test-retest reliability. In all types of analysis, the reliability level of STAR Reading 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 Reading's "content validity," that is, how the items are relevant to reading comprehension. 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 Reading results along with their scores on other assessments, such as the California Achievement Test, DIBELS, FCAT, Iowa Test of Basic Skills, and Stanford Achievement Test. Our analysis showed a correlation with these tests that exceeded the guideline provided by NCRTI. The technical manual provides details.

Summary

STAR READING BASICS

- STAR Reading is designed for students who have a sight vocabulary of at least 100 words.
- The test is typically administered in fall, winter, and spring for screening purposes and as often as weekly when monitoring the progress of students who are in intervention.
- The software adjusts the difficulty of each item to a student's performance. The test is administered in about 10 to 15 minutes.
- The software calculates various scores, including scaled scores, percentile ranks, instructional reading levels, and grade equivalents, which are used for different purposes.
- STAR Reading exceeds standards for reliability and validity.