Research Foundation for Star Adaptive Assessments
Science of Star
Reports and screens are regularly reviewed and may vary from those shown as enhancements are made.

All logos, designs, and brand names for Renaissance's products and services, including but not limited to Renaissance Growth Platform, Star Assessments, Star Early Literacy, Star Spanish Early Literacy, Star Math, Star Spanish Math, Star Reading, Star Spanish Reading, and Renaissance, are trademarks of Renaissance Learning, Inc., and its subsidiaries, registered, common law, or pending registration in the United States and other countries. All other product and company names should be considered the property of their respective companies and organizations.

© 2020 by Renaissance Learning, Inc. All rights reserved. Printed in the United States of America.

This publication is protected by U.S. and international copyright laws. It is unlawful to duplicate or reproduce any copyrighted material without authorization from the copyright holder. For more information, contact:

RENAISSANCE
P.O. Box 8036
Wisconsin Rapids, WI 54495-8036
(800) 338-4204
www.renaissance.com
educatordevelopment@renaissance.com

09/20
Dear Educators,

Our mission at Renaissance® “is to accelerate learning for all children and adults of all ability levels and ethnic and social backgrounds, worldwide.”

A key principle supporting this mission is that the first step in accelerating learning is to measure its occurrence. Renaissance is the leader in pre-K–12 learning analytics, offering solutions such as our computer adaptive Renaissance Star Assessments®, which enable teachers and curriculum creators to access immediate data on student learning and drive phenomenal student growth.

Students at tens of thousands of schools worldwide take assessments to measure their reading and math understanding using Renaissance Star Early Literacy®, Renaissance Star Reading®, and Renaissance Star Math® in English and Spanish. Educators like Star Assessments because they are easy to use, take relatively little time to administer, and provide teachers with highly valid and reliable data that meets the high review standards of the National Center on Intensive Intervention.

These multipurpose tools for screening and progress monitoring also include resources to target instruction for all kinds of learners. Upon completion of a Star test, students who are most at risk can be identified quickly and their needs diagnosed, so that intervention can begin immediately.

Read on and you will see why teachers using Star measures are able to accelerate learning, get more satisfaction from teaching, and help their students achieve higher scores on state and national tests. The stakes are high. We must help all students in schools everywhere be prepared for college or the workforce by the time they graduate. (Full technical manuals also are available for each Star assessment by request to research@renaissance.com.)

Sincerely,

James R. McBride, Ph.D.
Vice President & Chief Psychometrician
RENAISSANCE

James R. McBride, Ph.D., is vice president and chief psychometrician for Renaissance Learning. He was a leader of the pioneering work related to computerized adaptive testing (CAT) conducted by the Department of Defense. McBride has been instrumental in the practical application of item response theory (IRT) and since 1976 has conducted test development and personnel research for a variety of organizations. At Renaissance Learning, he has contributed to the psychometric research and development of Star Math, Star Reading, and Star Early Literacy. McBride is co-editor of a leading book on the development of CAT and has authored numerous journal articles, professional papers, book chapters, and technical reports.
Introduction

The importance of assessing students’ skill mastery early on in schooling is paramount. Research suggests that successful early intervention is the best single predictor of future academic success, particularly in critical areas like reading, language acquisition, and mathematics.

At their core, Star Assessments are purposeful, proven, powerful, and predictive (see figure 1). These computer adaptive assessments are designed to help teachers assess students quickly, accurately, and efficiently. Star Assessments provide teachers with reliable and valid data in an instant so they can target instruction, monitor progress, provide students with appropriate instructional materials, and intervene with at-risk students. Administrators use immediate feedback from these assessments to make decisions about curriculum, assessment, and instruction at the classroom, school, and district levels.

Figure 1. Star Assessments pillars

<table>
<thead>
<tr>
<th>Purposeful</th>
<th>Proven</th>
<th>Powerful</th>
<th>Predictive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data and insights educators need to address assessment questions and inform instructional decisions.</td>
<td>Trusted, reliable and valid data backed by research, validity studies, and millions of data points.</td>
<td>Cutting-edge learning science, data analytics, and test design delivering maximum impact in minimal time.</td>
<td>Statistical linking backing highly predictive data on student performance on state and other high-stakes tests.</td>
</tr>
</tbody>
</table>
Overview of Star Assessments

Star Assessments are purposeful assessments used by educators for many reasons. Using this assessment solution for early literacy, reading, or math, educators can answer important questions, such as Are my students mastering key skills? What do they need to learn next? How should I plan instruction? Are my students growing at an appropriate rate? Are students responding to intervention?, and set goals towards state-standards mastery. The tests offer repeatability and flexibility in administration and a wealth of data for educators:

- Teachers use assessment results to facilitate personalized instruction, including identifying students who most need remediation or enrichment.
- Principals access assessment data through reporting on performance at individual, class, building, and district levels.
- Administrators and assessment specialists analyze results to compare data across schools, grades, and special student populations.

Students take Star Assessments on individual computers, iPads, Chromebooks, and tablets 7” or greater. The software delivers multiple-choice items one by one, and a student selects answers with a mouse, keyboard, or touchscreen. After an assessment is completed, the software calculates the student’s score and reports data about results for individual students, classes, and grades at the school or district level.

**Star Early Literacy**

Students are expected to develop a variety of early literacy skills as they progress from kindergarten through early elementary on the path to becoming proficient readers. Students’ progress reflects both the home literacy environment and educational interventions. The Star Early Literacy assessment is a brief standards-based test (completed by students, on average, in less than 10 minutes) of early literacy (22 items) and numeracy (5 items) development that measures student performance throughout grades K–3, and provides teachers with valuable information about early skill acquisition along a continuum of expectations.\(^1\)

Star Early Literacy helps teachers intervene at the start of a student’s formal learning, which is especially helpful for kids who enter schooling lacking the literacy experiences and foundational skills that help to ensure early literacy and numeracy development takes root. The assessment helps teachers monitor progress based on the specific needs of each student and is especially helpful in identifying students at risk for later reading failure. Data from this test can be used for a variety of purposes such as goal setting, outcomes assessment, instructional planning, and intervention.

Star Early Literacy was designed to continually assess a child’s status at different stages throughout the early growth period, pulling items from a bank of more than 3,400 items (see figure 2 for an item example).

\(^1\) Students who have established a 100-sight-word vocabulary or reached the Probable Reader stage of literacy development in Star Early Literacy typically are ready to take a Star Reading assessment. Although Star Early Literacy was designed for grades K–3, it can be used with older students, such as struggling readers, nonreaders, special education students, or English learners.
Star Early Literacy differs from other early literacy assessments because it is:

1. Computer administered and uses graphics, audio instructions, and automatic dictation of instructions and test questions, so most children can test without teacher assistance.
2. Computer adaptive, with content and difficulty level of each test administration tailored to individual student performance.
3. Completed by students, on average, in less than 10 minutes, while correlating highly with many more time-intensive standardized measures of early literacy, reading, and other readiness skills.

Star Reading
The Star Reading assessment measures students’ performance on key reading skills for grades K–12 via a brief standards-based test of general reading achievement, administering 34 questions that students complete, on average, in less than 20 minutes. Star Reading yields valuable data about acquisition of skills along a continuum of literary expectations that serves three purposes for school and district administrators: (1) to give teachers quick and accurate estimates of students’ reading achievement levels, (2) to assess reading achievement relative to national norms, and (3) to provide a means for monitoring growth in a consistent manner longitudinally for all students. The Star Reading item bank numbers over 6,600 items (see sample item, figure 3).
To develop Star Reading, Renaissance conducted extensive research, consulting with reading and assessment experts, to arrive at the skills most appropriate for assessing reading development. Developers studied several publications, including the 2010 Common Core State Standards (CCSS); the Reading Framework for the 2009 National Assessment of Educational Progress; the National Council of Teachers of English (2006) *Principles of Adolescent Literacy Reform* policy brief; and the Alliance for Excellent Education’s (2004) *Reading Next* report.

**Star Math**

The Star Math assessment provides a reliable and valid method for measuring students’ mathematical abilities and progress towards goals in grades K–12, via a brief assessment of 34 questions that students complete, on average, in less than 25 minutes. Teachers, principals, assessment directors, and district-level administrators use this standards-based assessment of general math achievement for instructional planning, growth measurement, and program evaluation. At the student level, Star Math serves many purposes, including screening, formative assessment, progress monitoring, calculating growth, and outcomes assessment. By administering this assessment on a regular basis, teachers can effectively monitor progress and make appropriate adjustments to instruction. Research firmly supports progress monitoring, which has been successful in a variety of educational settings (Foegen et al., 2007; Harkin et al., 2016).

Students taking a Star Math test use blank work paper and pencils during the administration. For specific assessment items, the assessment provides an onscreen calculator and/or reference sheet. The item bank for Star Math comprises more than 6,200 items (see sample item, figure 4).
Star Spanish Assessments

Star Spanish Early Literacy, Star Spanish Reading, and Star Spanish Math Assessments are designed both for Spanish-speaking students in monolingual English programs (as a means to assess skills in their home language) and for dual-language programs (e.g., bilingual, dual-language immersion, two-way immersion).

There is significant benefit to assessing in Spanish students who speak Spanish as their home language or those who are members of a dual-language program. Paired with Star Assessments in English, teachers can determine skill mastery in both languages and leverage that information to make instructional decisions. Star Spanish Math assesses appropriate math skills (as does Star Math English) but with questions and answers written in Spanish. Star Spanish Early Literacy and Star Spanish Reading have been developed from our Spanish-specific learning progression, La Progresion de la lectura de Renaissance.

The progression is based on a variety of standard sets for Spanish. Skills have been developed to align to those standards by our internal Spanish language arts team, supported by expert consultation and review from native Spanish-speaking curriculum experts.

All item development starts from the skills in the learning progression. Test items have been written by our native Spanish-speaking content developers and reviewed by our internal Spanish editorial team. All items avoid regional biases or dialectical differences.

Administration and test frequency

The item banks for Star Spanish Assessments include enough items to allow for monthly test administration: Star Spanish Early Literacy contains more than 770 assessment items and takes students, on average, less than 10 minutes to complete, Star Spanish Reading includes more than 2,200 items and takes, on average, less than 21 minutes to complete, and Star Spanish Math has over 2,100 items and takes, on average, less than 23 minutes.
For students whose home language is Spanish, we suggest districts make the decision on whether or not to administer Star Spanish Assessments. There are many factors to consider, for example, the students’ background with Spanish instruction. If the decision is to administer both the English and Spanish versions, we suggest English learners take both test types close together—ideally within a few days of each other—within a single testing/screening window, so that results are comparable for similar points in time.

Scores and reporting
With the introduction of authentic Spanish Reading learning progressions, educators can view scores and generate reports, in both English and Spanish, to help identify which skills students already know and which they are likely ready to work on next. For educators using Star Spanish Math, scores and reports are consistent whether assessing in English or Spanish and are aligned to state-specific math learning progressions. To help inform teachers’ instructional decisions, Star Spanish Assessments report both scaled scores and domain scores. The assessments were normed using a large amount of nationwide data to provide norm-referenced percentile rank, normal curve equivalent, and grade equivalent scores.

Test Design

Computer adaptive tests
These computer adaptive tests (CATs) continually adjust the difficulty of a test administration by selecting assessment items based on a student’s performance on each previously answered item. CATs shorten testing time and spare students both frustration in receiving items that are too difficult and boredom in receiving items that are too easy. A well-designed CAT is often two or more times as efficient as a conventional test.

Decades of research have shown that CATs can be considerably more efficient than conventional tests, which present all students with the same test questions (e.g., Lord, 1980; McBride & Martin, 1983). As noted by Weiss (2004), “Early evidence of improved measurement precision (reliability) and validity (e.g., Johnson & Weiss, 1980; Kingsbury & Weiss, 1980) and large reductions in the number of items administered (typically 50% or more) without having an impact on the psychometric characteristics of test scores for CAT have been confirmed in a number of recent studies (e.g., Mardberg & Carlstedt, 1998; Moreno & Segall, 1997)” (pp. 77–78). (For more about the reliability and validity of Star Assessments, see detailed information in Psychometric Properties, p. 27.)

Research also suggests that CATs are a sound choice for progress monitoring student performance in Response to Intervention (RTI), “a process of providing high quality interventions that are matched to student need, and uses frequent progress monitoring of students’ response to interventions to assist in making important educational decisions” (Bray & Kehle, 2011, p. 616). Progress-monitoring feedback is key to RTI as it tells educators which interventions are helping students most. “Progress-monitoring measures must be frequent, sensitive to instructional change over a short period of time, predictive of overall success as measured by the benchmark assessment, and able to drive instructional decisions (Shapiro, 2012, p. 9) ... Star measures offer an important and potentially valuable contribution to RTI” (p. 20) in the following ways:

---

2 Star also offers Star Curriculum Based Measures (CBM), information for which is available from www.renaissance.com/products/star-cbm/.

RENAISSANCE ©Copyright 2020 Renaissance Learning, Inc. All rights reserved.
• **Frequency of administration**—Star Assessments provide flexibility, so educators can assess students as often as needed, whether that is three times per school year for screening, monthly to monitor student progress during the year with time to change a student’s growth trajectory, or as often as weekly to monitor intervention for kids in tiers 2 and 3 of an RTI/MTSS framework.

• **Sensitivity**—Star Math and Star Reading are highly rated by the National Center on Intensive Intervention (NCII) (2018a, 2018b) for “Sensitivity to Student Learning: Reliability of Slope.”

• **Predictive power**—Star Math and Star Reading meet criteria set by the NCII (2019a, 2019b) for “Classification Accuracy,” meaning their ability to correctly classify students as at-risk for important benchmarks such as state summative assessments and college entrance tests. (For more on Star Assessments’ predictive power, see Frequency and Purpose, p. 22, and Psychometric Properties, p. 27.)

• **Impact on instructional decisions**—Renaissance state-specific learning progressions and La Progresion de la lectura de Renaissance, work in tandem with Star Assessments, using scores from Star to place students within a continuum of learning that makes assessment data immediately actionable for instructional planning. With Renaissance learning progressions, educators know which topics students are ready to learn next and which critical Focus Skills™ they must master before tackling more advanced concepts. Notably, Star Math and Star Reading meet partial criteria set by NCII (2018a, 2018b) for “Decision Rules: Setting & Revising Goals.”

**Research support by the numbers**

Star Early Literacy, Star Reading, and Star Math assessments in English and Spanish are supported by extensive internal and external research (see table 1).

<table>
<thead>
<tr>
<th>Table 1. Research support for Star Assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star Early Literacy research</td>
</tr>
<tr>
<td>✓ 25 research pieces, 16 led independently</td>
</tr>
<tr>
<td>✓ 1 (Star Spanish Early Literacy)</td>
</tr>
<tr>
<td>Star Reading research</td>
</tr>
<tr>
<td>✓ 109 research pieces, 35 led independently</td>
</tr>
<tr>
<td>✓ 2 (Star Spanish Reading)</td>
</tr>
<tr>
<td>Star Math research</td>
</tr>
<tr>
<td>✓ 97 research pieces, 31 led independently</td>
</tr>
<tr>
<td>✓ 2 (Star Spanish Math)</td>
</tr>
</tbody>
</table>

**Item response theory and its role in CAT**

Tailoring item difficulty to match a student’s knowledge or skill level can be done in several ways; however, most CATs use item response theory (IRT) as the basis for both adaptive item selection and test scoring. IRT puts student performance and item difficulty on the same scale and offers a means to estimate the probability a student will answer a given test item correctly. IRT models provide a way to measure each item’s degree of difficulty and to estimate each student’s achievement level from a pattern of correct and incorrect responses to items.
Based on IRT, Renaissance reports topic-level scores for all Star Assessments, an approach highlighted by the Institute of Education Sciences (IES) in 2020 (Provasnik et al.). Renaissance first introduced “domain scoring,” nearly 20 years ago by reporting Domain and Skillset scores in Star Early Literacy in 2001. Domain scoring was then expanded to Star Math and Star Reading and is now the basis of both domain and mastery scores in all Star Assessments.

With IRT, scientists can calculate the probability of a correct response to an item as a function of student ability. As student ability increases, so does the probability the student will answer correctly. Additionally, because some test items are more difficult than others, the probability trend differs from one item to another. Figure 5 shows the probability functions for three test items: one item that’s easy, one that’s moderately difficult, and one that’s very difficult.

Figure 5. A student’s reaction to three test items of varying difficulty

During administration of a Star assessment, the software automatically moves up or down the item-difficulty scale to select questions based on a student’s previous answers. If the student answers a question correctly, the next question is usually more difficult. If the student answers incorrectly, the next question is generally less difficult. Unlike manual paper-and-pencil assessments, Star Assessments dynamically adjust to students’ unique responses to pinpoint student achievement levels quickly and efficiently.

Figure 6 illustrates a CAT administration, where items of varying difficulty are presented based on a student’s previous responses. This figure also shows how selecting items tailored to a student’s ability helps to reduce measurement error as the test progresses.
Multiple-choice format

In developing test items for Star Assessments, Renaissance researched and prototyped several item-response formats and ultimately chose to use multiple choice. Much research supports the use of multiple-choice items, also called selected-response format. As noted by Stiggins (2005),

> These tests are efficient in that we can administer large numbers of multiple-choice or true/false test items per unit of testing time. Thus, they permit us to sample widely and draw relatively confident generalizations from the content sampled. For this reason, when the target is knowledge mastery, selected-response formats fit nicely into the resource realities of most classrooms. (p. 70)

Multiple-choice items work well with computerized scoring, which automates the testing process and saves teachers time in collecting and scoring results (Nicol, 2007). Many multiple-choice test items can be administered in a short amount of time, which aids measurement precision. According to Haladyna and Downing (1989), “the use of multiple-choice formats generally leads to more content-valid test score interpretations.”

To construct multiple-choice items with a balanced range of cognitive complexity, Renaissance item specifications require: (1) verifying the accuracy of all content; (2) using grade-level appropriate cognitive load, vocabulary, syntax, and readability; (3) including only essential text and graphics; and (4) employing standards for bias, fairness, and sensitivity. Research has shown that well-designed multiple-choice questions can assess an array of skills (Cassels & Johnstone, 1984; Popham, 2008; Russell et al., 2003) at higher levels of student learning (Cox, 1976; Johnstone & Arbusaidi, 2000; Mattimore, 2009; Osterlind, 1998; Popham, 2003).
Test Content

Star Early Literacy, Star Reading, and Star Math assessments in English and Spanish are powerful, computer adaptive tests that save teachers time by automating administration and scoring and assessing students on a large and varied range of skills using relatively few items. Using a well-designed CAT, students spend less time spent completing the assessment—to equal the reliability of a 50-item conventional test, a well-designed CAT may use only 25 items to yield the same information in half the time.

Large item banks

Star Assessments have large item banks to allow multiple administrations without risk of item overexposure. Renaissance continuously develops high-quality assessment items to add to the banks to support frequent testing and achieve an even distribution of items across the difficulty levels of each Star assessment.

Star Assessments are fixed-length tests, which means item count is the sole criterion for ending an administration. For both English and Spanish, Star Early Literacy administers 27 items per test event, and Star Reading and Star Math each administer 34 items. The assessments were developed to provide precise measurement of student achievement in English and/or Spanish of early literacy (and early numeracy), reading, and math, and to do so efficiently.

Item-development process

Item development is of critical concern to Renaissance and has earned Star Assessments high ratings from the National Center on Intensive Intervention (see Test Design, p. 10, and Psychometric Properties, p. 27). Professional designers, writers, and editors—who have education backgrounds and content-area expertise—develop all content for Star Assessments. These experts follow research-based practices for developing assessment items and rigorously adhere to the following process to ensure quality item creation:

- Analyze standards to be assessed in the categories of skill, action, vocabulary, and context; refer to national or state resources for appropriate standard and grade-level expectation interpretation.
- Write item specifications and provide training to item writers and editors.
- Establish item metadata to guide development, including standards-related and item-related data.
- Use a multistep, recursive writing and editing process that ensures adherence to specifications as well as alignment to standards and item metadata.
- Post items for calibration and acquire student-response dynamic calibration data (more below).
- Examine psychometricians’ analyses of item-testing results.
- Add successful items to the operational assessment item bank.

Experts also receive ongoing item-writing training, which includes bias-and-fairness criteria to avoid stereotypes and characterizations of people or events that could be construed as demeaning, patronizing, or otherwise
insensitive. Content-development tools track and report attributes such as gender, age, ethnicity, subject matter, and regional references. Individual attributes, as well as the intersection of multiple attributes, are tracked throughout item development to ensure final content is demographically balanced and free of bias.

In addition, assessment items must pass strict quality reviews that check for discipline-specific criteria, accuracy, language appropriateness, and readability level; bias and fairness; and technical quality control.

Rules for item retention
Following these steps, information pertaining to each test item—including traditional- and IRT-analysis data, test level, form, and item identifier—is stored in an item-statistics database. A panel of content reviewers examines each item within content strands to determine whether the item meets all criteria for use in an operational assessment. After the content reviewers have designated items for elimination, recommendations are combined, and a second review is conducted to resolve any issues.

Dynamic calibration
To maintain and update the large item banks for each Star assessment, Renaissance continually develops and calibrates new test items using dynamic calibration. Each new assessment item is calibrated to determine its exact point on the Star difficulty scale.

For dynamic calibration, one or more new items are embedded at random points in a Star test and administered to large samples of students. Renaissance psychometricians collect student-response and other data on the item, and then perform a statistical analysis to pinpoint the new items’ scale values.

Items undergoing calibration do not count toward students’ scores on the Star assessment. On average, two or three additional items per test appear when calibration is turned on, and testing time is increased by approximately one minute. Norming, reliability, and validity studies take place after items successfully pass through calibration.

Learning Progressions–The Bridge Between Assessment and Instruction
Teachers using Star Assessments can monitor progress toward college- and career-readiness standards as well as predict proficiency on state tests. After a student completes a Star assessment, the software maps the resulting scaled score to an appropriate entry point in a state-specific learning progression, or La Progresion de la lectura de Renaissance, which then displays for educators the skills a student is likely ready to learn next.

A learning progression is a continuum of expected learning, beginning with emergent reading (or early numeracy) skills and progressing to the level of competence required for college and the workforce (see figure 7). The skills are interconnected and related, formed from requisites and prerequisites, and represent how students typically advance their learning in a subject area. According to Heritage (2008), “Learning progressions that clearly articulate
a progression of learning in a domain can provide the big picture of what is to be learned, support instructional planning, and act as a touchstone for formative assessment” (p. 1).

**Figure 7. Example learning progression for reading**

![Example learning progression for reading](image)

The Renaissance learning progression development teams follow a specific process (see figure 8) when creating learning progressions, including closely following best practices in standards research and alignment, as well as maintaining ongoing relationships in research and consultation with leading educational organizations, such as Mid-continent Research for Education and Learning (McREL) and the Northwest Regional Educational Laboratory (NWREL).

**Figure 8. A four-step process guided the development of Renaissance learning progressions**

<table>
<thead>
<tr>
<th>Step 1: Conduct Research</th>
<th>Step 2: Analyze Data</th>
<th>Step 3: Consult Experts</th>
<th>Step 4: Continued Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built initial set of skills based on survey of research literature, consultation with experts, and analysis of state and national standards.</td>
<td>Calibrated items to determine skill difficulty and to empirically validate order of skills.</td>
<td>Consulted experts including Dr. Margaret Heritage, Dr. Karin Hess, and Dr. James Milgram.</td>
<td>Ongoing annual consultation of experts and refreshing of data ensure precision.</td>
</tr>
</tbody>
</table>
The roadmap of skills presented in Renaissance learning progressions helps teachers monitor progress toward standards expectations and displays student progress relative to grade-level expectations. Because all students develop skills at different rates and in different ways, a student’s progression may not follow a straight trajectory. The Renaissance Growth Platform™ (see figure 9) links results from Star Assessments with learning progression information to create a personalized learning journey for each student, which includes grade-level domain expectations identified, resources for targeted instruction, intervention, and enrichment, as well as assets from third-party partners. Creation of additional content is continuously underway to help educators further pinpoint students’ knowledge and skills development.

**Figure 9. Learning progressions inform instruction via the Renaissance Growth Platform**

![Learning Progressions](image)

**Evolution of Renaissance learning progressions**

To create that link between assessment and instruction, the Renaissance learning progression development teams initially created two learning progressions, one for reading and one for math. The team rigorously developed, tested, and validated these progressions. For both subjects, educational content experts identified the initial order of item difficulty by researching reading and math theory, examining widely accepted frameworks and state standards, reviewing college- and career-readiness standards, and consulting nationally recognized subject-area experts.

*Standards alignment and reporting with individual state standards*
Then in 2013, as states continued to move toward more rigorous educational standards, the team became immersed in Common Core State Standards and other new state standards, as well the literature, resources, interpretations, and implementation information surrounding them. In 2013, Renaissance released reading and math progressions expressly for the Common Core, and by 2016, had created a state-specific learning progression for each state.

In creating the new learning progressions, Renaissance experts closely analyzed new state and national standards to identify each standard’s inherent skills, intent, and key terminology. They studied literature and resources regarding interpretation and implementation of the new standards by states and relevant consortia. Much time and effort ensured the new progressions included incremental steps of learning to fulfill the intent of the standards and ultimately culminate in students’ readiness for college and beyond (see figure 10). For in-depth information on the development of the learning progressions, see http://doc.renlearn.com/KMNet/R0053985FA6D567F.pdf.

Figure 10. Building state-specific learning progressions

Standards
Captured full hierarchy of standards and ancillary documentation to ensure full interpretation of standards.

Skills
Developed skill statements to represent intent of the standards.

Analysis
Identified skills inherent in standards. Studied horizontal and vertical alignment of standards in relation to other domains and grades.

Order
Placed skills in teachable order, reflecting how students learn and traversing domains as teaching naturally occurs. Validated using skill difficulty based on Star items.

Star State Standards Reporting (see figure 11) helps educators estimate a student, class, or district’s level of mastery on individual state standards (including CCSS). To develop this reporting, Renaissance used both empirical data and content-area expert review, similar to the method used by states to place their standards on state test scales. Standards were aligned with the Star scale using the following rigorous five-step approach:
1. Identify standards
2. Identify Star skills and items that assess the standard’s expectations for the skills and concepts.
3. Review the calibrated (research-based) difficulty level of Star items associated with the skills embedded in the standard.
4. Assign a difficulty level to the standard based on review of empirical data. Evaluate likely mastery for each standard, determined by its difficulty level and Star test scores.
5. Review assigned Star scaled score by content-area expert in relation to the composite standard to ensure placement is accurate and appropriate.

Figure 11. State Standards Report—Student, class, and district reporting available

La Progresión de la lectura de Renaissance
In 2019, the Renaissance Spanish Biliteracy Advisory Council was established to define a vision for how to close gaps in current educational equity. The council was made up of Spanish biliteracy experts, who identified the importance of a learning progression that would support Spanish-speaking students from early literacy through the end of high school.

Renaissance Spanish educational content and learning progression developers met this challenge using a process of standards and literature research; cycles of expert consultation, development, and expert review; and validation via Star Spanish Reading assessment item and skill difficulty data.

The developers researched Spanish language acquisition literature and analyzed key Spanish Reading standards: California CCSS en Español, CCSS en Español, TEKS Spanish Language Arts and Reading, Estándares de Contenido y Expectativas de Grado de Puerto Rico (Puerto Rico Core Standards), Aprendizajes clave para la educación integral: Plan y programas de estudio para la educación básica (México), Currículum nacional de Chile, and Currículo nacional de la educación básica (Perú).
The result was the Progresión de la lectura de Renaissance, a coherent and continuous pathway on which Spanish-speaking students move incrementally through levels of increasing competence in Spanish Reading. Every incremental level further builds upon and integrates the previous one as students accumulate expertise on a path to achieving the Spanish Reading goals set by state and agency standards.

Each skill in the continuum is placed when it could sensibly be taught in relation to all other skills intended for that year. Skills were first ordered based on how students would typically be expected to learn them, and then subsequently validated by applying student data from thousands of Spanish assessment items. Like all Renaissance learning progressions, the suggested sequence provides context for where students are in their learning and to answer the question "what comes next?"

**Path from test blueprint to learning progression**

Empirical testing has continued to find a strong statistical link between the progression of skills in the learning progressions and the assessed difficulty level of Star assessment items. Figure 12 depicts the path from the Star test blueprint to a student’s placement in the learning progressions.

The Star test blueprint (left side of figure) shapes each student's test, by working with CAT technology to determine which items are presented to the student. While each Star test event is unique, the blueprint ensures a certain number of items from the domains and skill sets is included in each assessment. After the student has completed the assessment, the software uses the student’s resulting Star score to find an appropriate entry point in the learning progression (far right of figure), which then helps to inform next steps for the teacher and student.

**Instructional planning**

Learning progressions serve as a map for teachers to understand where students have been and where they are headed next on the path to readiness for college and the workforce. After a student has completed a Star assessment, their scaled score pinpoints an appropriate entry point in the learning progression, which displays the skills the student has likely mastered, the ones they are ready to learn next, and suggestions for the teacher to focus instruction.

Reporting in Star also provides teachers with lists of skills a whole class or individual students are ready to develop next (see figure 13), and teachers can access materials for the skills and domains reported on to determine next steps for students.
Focus skills

To help educators determine the most essential skills for reading and mathematics, Renaissance has identified Focus Skills in the learning progressions. Focus skills are the critical skills that students need to develop to move on to more advanced concepts. If time and other constraints prohibit educators from focusing deeply on each individual skill, Focus Skills in Star Assessments will help them determine the most important skills to target (see figure 14). For more information, see https://www.renaissance.com/focus-skills/.
Frequency and Purpose

Most schools administer Star Assessments to all students in the fall, winter, and spring for screening purposes. To establish a trend line for students (visible in Star reporting) and forecast proficiency on state tests or standards mastery, schools also administer an extra test in late fall. This way, after the winter screening, three data points have been established, which will allow the software to chart students’ growth trajectories.

Teachers who monitor progress more closely for specific students in intervention or other settings, or for instructional planning, typically test more frequently. Although Star Assessments can be administered as often as weekly, it is important to administer assessments to students only when a plan is in place to act upon the results. Teachers can use Star Assessments to capture each student’s overall growth from the beginning to the end of the school year, or in semester increments.
Growth measurement: Scaled score and student growth percentile

Because changes in student achievement do not happen overnight, measuring growth is essential to understanding the effects of instruction. Just as meteorologists use statistical models to predict the weather, educational researchers use growth models to identify patterns in student growth. Renaissance has unique insight into how students grow through ongoing study of Star data from millions of tests taken by students at thousands of schools and can approximate how much growth is typical for students of varying achievement levels in different grades from one time period to another.

Scaled score
Star Assessments generate a scaled score (SS), which is useful for comparing student performance over time. Within each assessment, the same range is used for all students, so scaled scores help to compare student performance across grade levels. Any scaled score increase indicates that a student has experienced growth. In 2017, updated norms for Star Early Literacy, Star Reading, and Star Math were developed using the Unified scale to provide educators with the most up-to-date picture of student performance. The Unified scale provides even greater measurement precision and one consistent scale across all computer-adaptive Star Assessments. Additionally, the Star Unified Scale connects Star Early Literacy and Star Reading assessment results so that educators can measure student achievement and growth on the same scale as students transition between the two tests.

Student growth percentile
Student growth percentile (SGP) compares a student’s growth to that of academic peers nationwide to help educators understand how students grow. A student’s academic peers are students at the same grade level with a similar score history. SGP was first developed by Dr. Damian Betebenner from the Center for Assessment, in partnership with measurement experts from several state departments of education.

An advantage of SGP is that it provides a clear picture of whether a student’s growth was more or less than expected. An increase in scores from one test administration to other is an indicator of growth; however, it does not indicate the type of growth. Is this growth low, typical, high? What is good growth? SGPs help educators understand, given where a student started, to what extent the growth achieved was as expected. For example, SGP 50 is median growth, meaning the student is on track with the peer group. SGP 25 is toward the low end of typical growth within the peer group and SGP 75 is toward the high end. So, for example, an SGP of 70 means a student grew as much or more than 70 percent of academic peers in the same grade with a similar prior score history. A student must take at least two Star Assessments within specific testing windows (fall to spring, fall to winter, winter to spring, spring to spring, spring to fall, or fall to fall) to generate an SGP score. (For more information on SGPs, see http://doc.renlearn.com/KMNet/R00571375CF36BBF.pdf.)

As more students test with Star Assessments, SGPs are periodically updated to ensure that the growth norms reflect the most up-to-date data possible. Currently, the Star Early Literacy dataset contains more than 3.7 million
students, Star Reading data includes over 30.9 million students, and Star Math data numbers 17.9 million students. The datasets for Star Spanish Assessments are as follows: Star Spanish Early Literacy—over 160,000 students; Star Spanish Reading—over 375,000 students; and Star Spanish Math—over 235,000 students. Using this data, Star SGPs drive goal setting in Star Assessments (see figure 15), which helps educators set challenging but reasonable, progress-monitoring goals personalized to each student.

Figure 15. Goal setting with Star Assessments

SGPs also appear in Star reporting to provide key information about student growth over time. SGP helps educators at classroom, school, and district levels address important questions about student progress using Star reporting tools, such as the Star Growth Report (see figure 16).
For teachers, Star reports help answer questions such as, *Did students grow from one testing period to the next? Did students of all abilities grow? Did students grow as much as expected? Did students in intervention grow more than their peers nationwide? Did my intervention strategies lead to greater growth?*

Likewise, administrators can use reporting to understand how much growth was achieved by all students in the district, whether students of all abilities grew, and if students grew as much as expected, as well as answer questions such as, *Did students in certain classes, grades, or schools grow more than others? What does overall student growth say about our core curriculum, intervention strategies, and programs and/or professional development needs?*
Star Assessments are predictive: Statistical linking studies, performance reporting

Star Assessments are reliable and valid and trusted by educators for a variety of purposes, including screening, progress monitoring, and intervention.

*Will my students perform well on the state test?* is one of the most serious and challenging questions that teachers and administrators face. Information from Star Assessments can help educators to evaluate all students’ progress toward proficiency.

Because Star Assessments are taken online, achievement data for millions of students nationwide is collected each year. Renaissance continuously conducts studies between the Star measures and various state and other assessments to statistically link the tests. For a comprehensive list of the assessments studied, see [http://doc.renlearn.com/KMNet/R0053964E87BAEC5.pdf](http://doc.renlearn.com/KMNet/R0053964E87BAEC5.pdf).

Renaissance’s linking studies drive the information displayed in Star State Performance reporting, which shows which students are meeting state proficiency requirements and which students may need additional support. Star Assessments monitor proficiency periodically—and more importantly, early—so that educators know whether students are on track to achieve proficiency on the state test. If not, there is still time to pivot instruction while there is still time to impact student performance (see figure 17).

**Figure 17. Star State Performance Reports—Student, class, and district reporting available**

![Pathway to Proficiency, Grade 6 Reading](image)

In addition to monitoring student proficiency, Star Assessments are approved for many uses and meet high technical standards.
Response to Intervention screening and progress monitoring

Response to Intervention—part of a Multi-Tier System of Supports (MTSS)—is a framework for making instructional decisions based on data, to accelerate learning for all students. Interim assessments, like Star Assessments, play a key role in RTI, by providing data to inform and improve instruction through screening/benchmarking and/or progress monitoring:

- As screening and benchmarking periodic assessments, typically administered two to four times per year to monitor growth of a group toward a proficiency target, which also provide information about standards students have likely mastered.
- As progress-monitoring assessments, administered more frequently than annually—as often as monthly to monitor students’ growth trajectories or weekly in intervention settings to measure individual progress. Progress-monitoring assessments measure growth during the year and longitudinally over two or more years. Also included in this category are diagnostic assessments administered as needed to help identify specific areas of weakness.

High-stakes purposes

Educators use assessments in a variety of ways. Some tests are used for a single purpose, while others, like Star Assessments, meet various needs. Star measures have many instructional uses, such as helping teachers determine what students know and are ready to learn next, how much they are growing, and whether they respond adequately to instruction. In addition, Star Assessments are approved by many states and districts for higher stakes decisions, as one of multiple data points considered for purposes including, but not limited to, the following:

- Student growth indicator in educator evaluation formulas
- Grade promotion
- Gifted & Talented identification
- Screening for characteristics of dyslexia

In addition, a growing number of research professors and educational research firms use Star Assessments for high-stakes purposes, such as serving as outcome measures to determine whether various interventions had the desired impact on student achievement and/or growth.

Psychometric Properties

Using Star Assessments for various purposes depends on specific state and district policies, but one commonality is that any assessments used show evidence of strong technical adequacy, including reliability, validity, and predictive accuracy. Star Assessments and Star Spanish Assessments meet these requirements.

Star Assessments have a long history of high ratings for reliability and validity

Since the first review of progress-monitoring tools in 2012, Star Assessments have been highly rated by the US Department of Education’s National Center on Intensive Intervention, whose mission is “to build state and district capacity to support educators in using data-based individualization to effectively implement intensive interventions.
in reading, mathematics, and behavior in Grades K–12” (https://intensiveintervention.org/). NCII has rated Star Assessments highly for both screening (2019a, 2019b) and progress monitoring (2016, 2018a, 2018b).

Prior to NCII, in 2009, the federally funded National Center on Response to Intervention (NCRTI), whose mission was “to provide technical assistance to states and districts and building the capacity of states to assist districts in implementing proven models for RTI/EIS” (now https://mtss4success.org/), listed Star Early Literacy, Star Reading, and Star Math among the first assessments designated as highly rated for screening and progress monitoring, ratings they maintained through years of subsequent reviews. And it was back in 2006, when Star Assessments received their first high marks as tools for RTI when the NCRTI’s predecessor, the National Center on Student Progress Monitoring, defined the assessments as reliable and valid for progress monitoring.

What is test reliability?
Test reliability is a measure of the consistency of test scores. Tests must yield consistent results to be useful. Two kinds of consistency inform an evaluation of a test’s measurement precision: (1) internal consistency, and (2) the consistency of scores obtained when an assessment is given two or more times.

Generic reliability, which is used to calculate internal consistency, applies conditional measurement error from individual students’ tests to estimate the percentage of score variation that is attributable to the feature the test is intended to measure. Retest reliability measures the consistency of scores across multiple administrations of an assessment to the same students, which is presented as a correlation coefficient between pairs of test scores earned by the same students on different occasions.

The quality (reliability) of the data provided to teachers is critical. A reliability coefficient of 1.0 = perfect (although perfect reliability is only theoretical), with coefficients of 0.7, 0.8, and 0.9 defined as good, better, and best. Figures 18 and 19 show the high reliability of Star Assessments, both in English and Spanish.

Figure 18. Star Assessments rate highly in reliability
What is test validity?

Test validity was long described as the degree to which a test measures what it is intended to measure. A more current description is that a test is valid to the extent that there are evidentiary data to support specific claims as to what the test measures, the interpretation of its scores, and the uses for which it is recommended or applied. Evidence of test validity is often indirect and incremental, consisting of a variety of data that in the aggregate are consistent with the theory that the test measures the intended construct(s), or is suitable for its intended uses and interpretations of its scores. Determining the validity of a test involves the use of data and other information both internal and external to the test instrument itself.

Decades of research prove the validity of Star Assessments (see figure 20).
Evidence of psychometric reliability combined with high alignment of test content to curriculum standards may be evidence enough of an assessment's validity; however, other measures further strengthen its validity: cumulative evidence of criterion-related validity, convergent and discriminant validity evidence, and demonstrated accuracy of screening and diagnostic classifications, among others. These components of construct validity provide evidence that an assessment measures specific attributes as claimed and is appropriate for specific uses and inferences. Construct validity evidence is attained cumulatively; upon its initial release, an assessment may have evidence consistent with construct validity, but over time additional support should be accumulated and documented.

**Technical Manuals**

For more correlations and additional information, full technical manuals are available upon request for Star Assessments at research@renaissance.com.
References


Acknowledgements

The following experts have advised Renaissance in the development of Star Assessments.

Catherine N. Close, Ph.D., is director of psychometrics for Renaissance with extensive background in both classical and modern test theories and their application to computerized adaptive tests (CAT) and conventional non-adaptive tests.

Margaret Heritage, Ph.D., is assistant director for professional development at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA, working on the development of academic language for EL students and formative assessment, including teachers’ use of formative assessment evidence.

Karin Hess, Ed.D., a senior associate with the Center for Assessment (NCIEA) since 2002, is recognized nationally for her research and work with learning progressions, text complexity, performance assessment, and cognitive rigor.

Thomas P. Hogan, Ph.D., is a professor of psychology and a Distinguished University Fellow at the University of Scranton with more than 40 years of experience conducting reviews of mathematics curricular content, principally in connection with the preparation of a wide variety of educational tests.

Xiaoying Ma, Ed.D., is senior psychometrician for Renaissance with extensive background in language testing (TOEFL) and large-scale state testing programs.

James R. McBride, Ph.D., is vice president and chief psychometrician for Renaissance, and was a leader of the pioneering work related to computerized adaptive testing (CAT), including that conducted by the Department of Defense, and development of the earliest commercial applications of CAT.

R. James Milgram, Ph.D., is a professor of mathematics at Stanford University, whose work in mathematics education includes consulting with several states on math standards, including California.

Michael Milone, Ph.D., is a research psychologist and an award-winning educational writer and consultant to publishers and school districts.

James B. Olsen, Ph.D., is a psychometrician for Renaissance, with professional skills in educational and professional assessment, job analysis, test design and development, psychometric and statistical analysis, research and evaluation, and professional certification and accreditation.

Adam E. Wyse, Ph.D., is a senior psychometrician for Renaissance with vast expertise in assessment and psychometrics who has published many research articles related to these areas in well-known scholarly journals.

James Ysseldyke, Ph.D., is Emma Birkmaier Professor of Educational Leadership in the Department of Educational Psychology at the University of Minnesota, educating school psychologists and researchers for more than 35 years.
Acknowledgements

The following experts have advised Renaissance in the development of Star Assessments.

Catherine N. Close, Ph.D.,
director of psychometrics for Renaissance.

R. James Milgram, Ph.D.,
professor of mathematics at Stanford University.

Margaret Heritage, Ph.D.,
assistant director for professional development at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), UCLA.

Michael Milone, Ph.D.,
research psychologist and award-winning educational writer and consultant to publishers and school districts.

Karin Hess, Ed.D.,
senior associate with the Center for Assessment (NCIEA).

James B. Olsen, Ph.D.,
psychometrician for Renaissance.

Thomas P. Hogan, Ph.D.,
professor of psychology and a Distinguished University Fellow at the University of Scranton.

Adam E. Wyse, Ph.D.,
senior psychometrician for Renaissance.

Xiaoying Ma, Ed.D.,
senior psychometrician for Renaissance.

James Ysseldyke, Ph.D.,
Emma Birkmaier Professor of Educational Leadership in the Department of Educational Psychology at the University of Minnesota.

James R. McBride, Ph.D.,
vice president and chief psychometrician for Renaissance.