Renaissance

Virginia Growth Assessments: Alternative Assessment Submission Form

FINAL SUBMISSION

Virginia Department of Education
Attn: Katie Carroll
101 N. 14th Street, 21st Floor
Richmond, VA 23219

Cover Letter

Renaissance
See Every Student.

PO Box 8036
Wisconsin Rapids, WI
54495-8036

P: (715) 424-3636
www.renaissance.com

June 26, 2024

Virginia Department of Education
Katie Carroll, Deputy Superintendent of School Performance and Support
101 N. 14th Street, 21st Floor
Richmond, VA 23219

RE: Alternative Assessment Submission Form

Dear Ms. Carroll:

Renaissance Learning is pleased to submit to the Virginia Department of Education (VDOE) the Virginia Growth Assessments: Alternative Assessment Submission Form for the use of Star Reading and Star Math as an alternative assessment to the Virginia Growth Assessment during the 2024-2026 school years. The form, along with corresponding documentation in this submission, demonstrates Star Reading and Star Math alignment to the Standards of Learning (SOL) and meets the requirements of Senate Bill 345 and House Bill 1076.

Since 2017, Renaissance has proudly served Virginia Divisions using Star Reading and Star Math as their alternative growth assessment. Built on sound psychometrics and years of research, Star assessments measure student growth and help educators provide targeted instruction to improve student achievement.

We look forward to continuing our partnership with the VDOE and divisions to support student academic achievement and growth. For more information on this proposal please contact Michael Hurst, Vice President of Proposal Solutions and State Partnerships, at (800) 338-4204 or proposals@renaissance.com for more information.

Sincerely,

Ted Wolf

Ted Wolf
Senior Vice President and Global Controller
Renaissance Learning, Inc.

Documentation Specifications

Documentation Specifications

Requirement Robust documentation demonstrating alignment with the Standards of Learning

As reflected in the Star Math Alignment and Star Reading Alignment reports included in the Documentation Section (pages 12-88 of the submission), the Star Math and Star Reading items are aligned to the 2023 Math and 2024 Reading Virginia Standards of Learning (SOL). Behind our more than 38 years of experience researching and analyzing state standards, and aligning practice and assessment content, our alignment services team ensures teachers can accurately measure student growth in terms of standards, core skills, and objectives, and adjust instruction accordingly.

To ensure that each item-to-standard relationship is precise and dependable, Renaissance uses a clear and rigorous alignment strategy. This strategy was developed and refined through consultations with Education Northwest and McREL. Experienced education specialists meticulously analyze skills within the standards and the product to produce accurate alignments based on defined best practices that include the following steps:

- Validate official state standard documentation, assessments, and crosswalks.
- Analyze both the content and cognitive demand of the standards.
- Examine the full intent and scope of each standards statement, considering the skill components and the standards hierarchy.
- Identify the skills by looking at specific skill phrases, content vocabulary, academic vocabulary, condition and boundaries, and characteristics related to Webb's Depth of Knowledge.

Our standards analysis process and skill alignment process ensure that the product skill or item does not extend beyond the domain and intent of the standard. In addition, it ensures that when a student works through an item, the student is genuinely assessed on the expectation for which that item has been developed.

Requirement Technical report documenting validity and reliability of the alternative assessment

Evidence of reliability and validity for both Star Math and Star Reading assessments can be found in the **Star Math and Star Reading Technical Manuals** included in the Documentation Section of this submission (pages 102-494) . Specifically, Star Reading assessment reliability and validity information can be found on pages 48-57 of the **Star Reading Technical Manual** and pages 64-95 of the Star Reading Technical Manual. Pages 40-53 and pages 54-79 of the **Star Math Technical Manua**l discuss the reliability and validity of the Star Math assessment.

In addition to the Star Reading and Math Technical Manuals, the National Center on Intensive Intervention determined Star Math and Star Reading are highly effective tools for both academic screening https://charts.intensiveintervention.org/ascreening and progress monitoring https://charts.intensiveintervention.org/aprogressmonitoring. The independent analysis evaluated the technical rigor of the tools, including classification accuracy, technical standards, and usability features.

Requirement	Documentation that alternative assessment(s) includes at least one beginning-of-year assessment, one mid-year assessment, and one end-of-year assessment.

Both **Star Reading** and **Star Math** can be administered three times per year (**beginning-**, **mid-**, and **end-of-year**). This frequent administration provides both an absolute and normed growth measure. The Renaissance Star Implementation Guide included in the **Documentation Section** (pages 90-101 of the submission) offers best practices for **Star Assessment** implementation, including recommending testing three times a year.

Requirement	Technical report documenting the ability of the assessment to administer off-grade, on-grade, and above-grade items.

Star Reading and **Star Math** assessments assess student performance at, **above-**, and/or **below grade levels**. The test blueprints - covering a range of skills and skillsets aligned to Virginia's state-specific learning progression skills - include items of varying difficulty and complexity to ensure student performance is accurately assessed.

The **Star Reading Technical Manual** describes our test design approach in greater detail, including the **Adaptive Branching approach** (page 7); Content Specifications (pages 15-18); S**tar Reading and Learning Progressions for Reading** (pages 27-30); and **Computer-Adaptive Test Design** (pages 43-44).

Similarly, the **Star Math Technical Manual** describes our test design approach in greater detail, including the **Adaptive Branching approach** (page 6); **Content Specifications** (pages 13-15); **Star Math** and **Learning Progressions for Math** (pages 21-23); and **Computer-Adaptive Test Design** (pages 36-37).

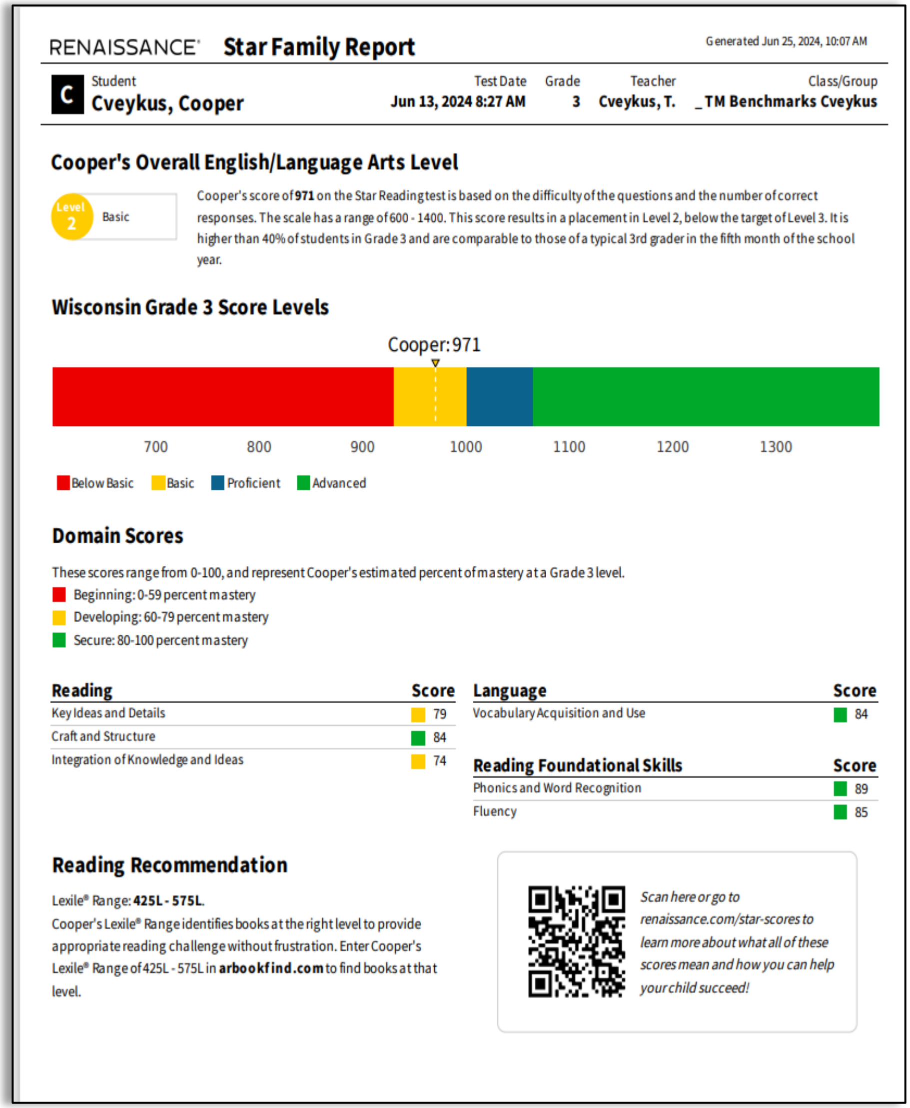
Requirement	Technical report documenting the ability to report individual student growth scores over the course of the school year.

Pages 113-114 of the **Star Reading Technical Manual** document the ability to report individual student growth scores over the course of the school year.

Pages 87-89 of the **Star Math Technical Manual** document the ability to report individual student growth scores over the course of the school year.

Requirement Example of the parent/family report and when it will be available to school divisions.

Star assessments include a **Star Family Report** (see example on right). Reports are available immediately upon completed testing. The Star Family Report serves as a bridge between a student's academic home life and classroom experience, and it helps parents and guardians monitor a student's progress. **Available in both English and Spanish, the Star Family Report summarizes a student's test results and explains what the scores mean.**



RENAISSANCE **Star Family Report**

Generated Jun 25, 2024, 10:07 AM

Student	Test Date	Grade	Teacher	Class/Group
C **Cveykus, Cooper**	**Jun 13, 2024 8:27 AM**	**3**	**Cveykus, T.**	**_TM Benchmarks Cveykus**

Cooper's Overall English/Language Arts Level

Level 2 Basic

Cooper's score of **971** on the Star Reading test is based on the difficulty of the questions and the number of correct responses. The scale has a range of 600 - 1400. This score results in a placement in Level 2, below the target of Level 3. It is higher than 40% of students in Grade 3 and are comparable to those of a typical 3rd grader in the fifth month of the school year.

Wisconsin Grade 3 Score Levels

Domain Scores

These scores range from 0-100, and represent Cooper's estimated percent of mastery at a Grade 3 level.

- Beginning: 0-59 percent mastery
- Developing: 60-79 percent mastery
- Secure: 80-100 percent mastery

Reading	Score
Key Ideas and Details	79
Craft and Structure	84
Integration of Knowledge and Ideas	74

Language	Score
Vocabulary Acquisition and Use	84

Reading Foundational Skills	Score
Phonics and Word Recognition	89
Fluency	85

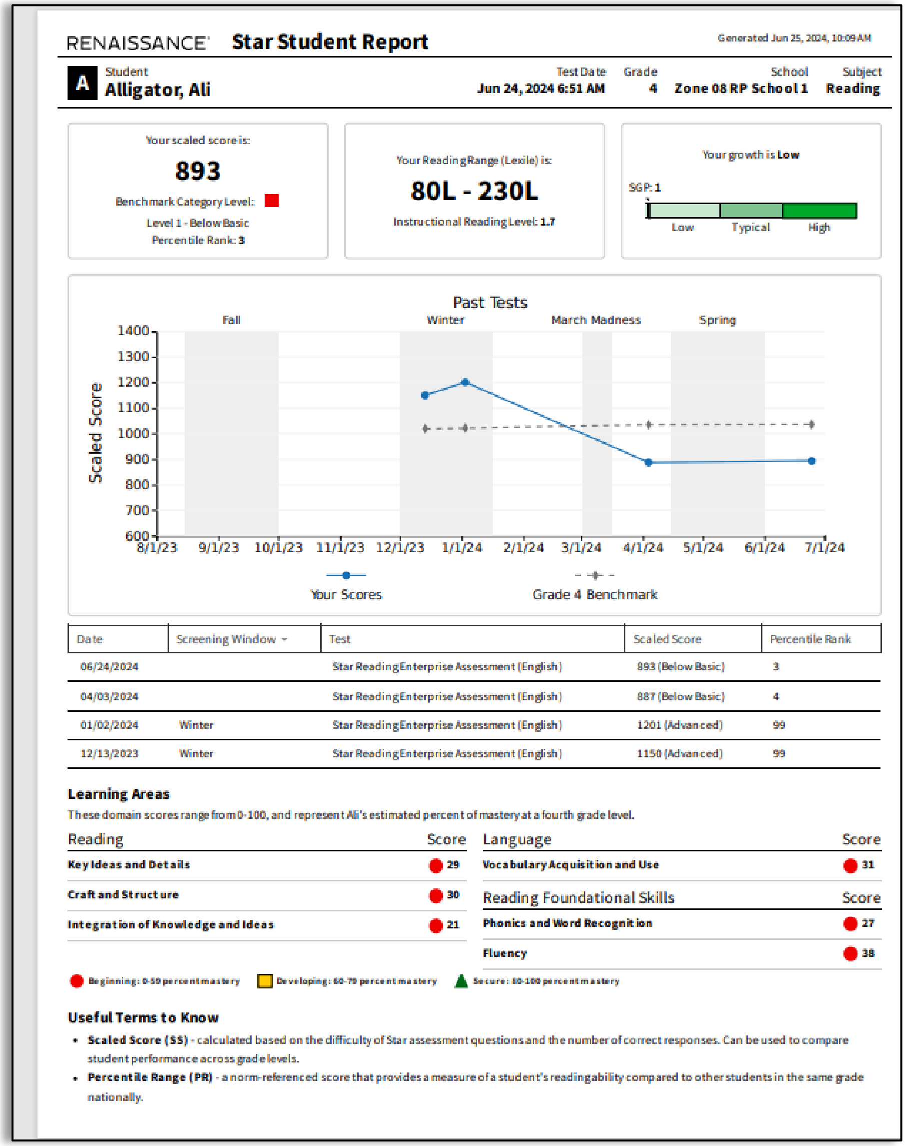
Reading Recommendation

Lexile® Range: **425L - 575L**.

Cooper's Lexile® Range identifies books at the right level to provide appropriate reading challenge without frustration. Enter Cooper's Lexile® Range of 425L - 575L in **arbookfind.com** to find books at that level.

Scan here or go to renaissance.com/star-scores to learn more about what all of these scores mean and how you can help your child succeed!

Other student- specific reports, such as the **Star Student Report** (see example below) can also be printed to provide additional information for parents/families as needed.



RENAISSANCE **Star Student Report** Generated Jun 25, 2024, 10:09 AM

Student	Test Date	Grade	School	Subject
Alligator, Ali	**Jun 24, 2024 6:51 AM**	**4**	**Zone 08 RP School 1**	**Reading**

Your scaled score is: **893**
Benchmark Category Level: Level 1 - Below Basic
Percentile Rank: **3**

Your Reading Range (Lexile) is: **80L - 230L**
Instructional Reading Level: **1.7**

Your growth is **Low**
SGP: **1**
Low Typical High

Date	Screening Window	Test	Scaled Score	Percentile Rank
06/24/2024		Star Reading Enterprise Assessment (English)	893 (Below Basic)	3
04/03/2024		Star Reading Enterprise Assessment (English)	887 (Below Basic)	4
01/02/2024	Winter	Star Reading Enterprise Assessment (English)	1201 (Advanced)	99
12/13/2023	Winter	Star Reading Enterprise Assessment (English)	1150 (Advanced)	99

Learning Areas

These domain scores range from 0-100, and represent Ali's estimated percent of mastery at a fourth grade level.

Reading	Score
Key Ideas and Details	29
Craft and Structure	30
Integration of Knowledge and Ideas	21

Language	Score
Vocabulary Acquisition and Use	31

Reading Foundational Skills	Score
Phonics and Word Recognition	27
Fluency	38

Beginning: 0-59 percent mastery; Developing: 60-79 percent mastery; Secure: 80-100 percent mastery

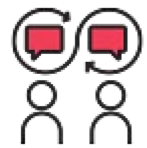
Useful Terms to Know

- **Scaled Score (SS)** - calculated based on the difficulty of Star assessment questions and the number of correct responses. Can be used to compare student performance across grade levels.
- **Percentile Range (PR)** - a norm-referenced score that provides a measure of a student's reading ability compared to other students in the same grade nationally.

Requirement	List of training modules for teachers and principals on interpretation and use of student growth data.

Renaissance offers concise and thorough learning opportunities that are based on **face-to-face onsite training**, **virtual webinars**, **recorded online learning**, **on-demand learning modules**, and **monthly office hours**.

The following training modules are available for teachers and principals to support the interpretation and use of **Star Reading** and **Star Math** student growth data.



Smart Start Online Learning Modules for Star Reading and Star Math

Smart Start product training is a **self-paced**, **on-demand**, **multi-media program** that outlines critical first steps to start using **Star Reading** and **Star Math**. Most users can complete all Smart Start training segments for Star assessments in **45 minutes**. Content is chunked into brief, easy-to-understand segments. Smart Start provides just-in-time training.

Content is organized around the following learning outcomes:

- Planning for Star Assessments
- Overview of Star Assessments data
- Testing with fidelity
- A first look at Star data
- Establishing a good testing environment
- The importance of grouping
- From scores to skills to resources
- Using instructional planning reports
- Viewing results after testing

Train-the-trainer Learning Model for Leaders, Instructional Coaches, and Teachers

Professional Development	Session Outcomes
Topic: Understanding Star Reading and Star Math **Duration:** One 6-hour session **Delivery mode:** Face-to-face, train-the-trainer learning model **Location:** Onsite at 6 different educational centers around the state **Participants:** Leaders, Instructional Coaches, and Lead Teachers	Participants who complete this training session will understand their roles in the state implementation, as well as the purpose of the assessments. They also will learn to: • Create a plan for re-delivery of content • Understand how to give Star assessments with fidelity • Navigate the Star dashboard • Understand all components of Star Reading and Star Math • Comprehend key milestones and communication pathways, including daily information, product-specific questions, service escalations, and rostering concerns • Use goal settings for progress monitoring • Understanding how to screen for characteristics of dyslexia • Understand assessment metrics, including scaled scores, literacy classifications, and percentile rank • Understanding Parent and Family Reports, Data Interpretation

Remote Webinars

Remote Webinars are **60-minute live sessions** delivered by an expert consultant that focuses on a specific topic. The targeted audience will be teacher leaders and classroom teachers. Additional resources are provided with sessions to extend learning.

Webinar sessions offered:

- **How to Get Started Star Reading and Star Math Assessments** (5 sessions offered on different days and times)
- **How to Interpret Star Reading and Math Data** (5 sessions offered on different days and times)

- **How to Set Goals for Progress Monitoring** (5 sessions offered on different days and times)
- **How to Get Started with Reading CBM** (5 sessions offered on different days and times)
- **How to Screen for Characteristics of Dyslexia Using Star Data** (5 sessions offered on different days and times)
- **Decision-making for MTSS with Star Data** (5 sessions offered on different days and times)
- **How-to-webinars** (recorded)

Webinar courses include:

- **Getting Started with Star Assessment** – discover **Star Reading** and **Star Math** and how to administer and score the screener assessments with fidelity.
- **Interpreting Star Reading and Star Math** – explore **Star Early Literacy** and **Star Reading** screening reports and how the data informs instructional next steps.
- **Enhancing Instruction with Star Reading and Star Math** – gain further insight into student performance and instructional needs with the Star diagnostic assessments.

Office Hours: Ongoing Yearlong Learning Opportunities

Renaissance proposes **monthly online office hours** to support Virginia stakeholders including district-level administrators, test coordinators, school leaders, and teachers. During each 60-minute event, a Renaissance consultant will provide personalized support, answer questions, and offer additional learning opportunities focused on test administration, reporting, and interpreting screening results.

Office Hours	Session Outcomes
Topic: Office Hours **Timeframe:** Monthly, ongoing support **Duration:** 60-minute sessions **Delivery mode:** Virtual **Participants:** All stakeholders	Office Hours provide participants with monthly opportunities for continuous support of their Star assessment implementation. Such topics will include: • Use Star data to make instructional decisions • Understand how to incorporate Star computer-adaptive assessment into their intervention practices • Data interpretation • Goal setting for progress monitoring • Screening for characteristics of Dyslexia

Supporting Documentation

Star Math and Star Reading Alignment to Virginia's Standards of Learning

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
K	SOL.Math.CE.K.6	The student will model and solve single-step story and picture problems with sums to 10 and differences within 10, using concrete objects.	×
K	SOL.Math.MG.K.7	The student will recognize the attributes of a penny, nickel, dime, and quarter and identify the number of pennies equivalent to a nickel, a dime, and a quarter.	×
K	SOL.Math.MG.K.8	The student will investigate the passage of time by reading and interpreting a calendar.	×
K	SOL.Math.MG.K.9	The student will compare two objects or events, using direct comparisons, according to one or more of the following attributes: length (longer, shorter), height (taller, shorter), weight (heavier, lighter), temperature (hotter, colder), volume (more, less), and time (longer, shorter).	×
K	SOL.Math.MG.K.10.a	identify and describe plane figures (circle, triangle, square, and rectangle);	×
K	SOL.Math.MG.K.10.b	compare the size (smaller, larger) and shape of plane figures (circle, triangle, square, and rectangle); and	×
K	SOL.Math.MG.K.10.c	describe the location of one object relative to another (above, below, next to) and identify representations of plane figures (circle, triangle, square, and rectangle) regardless of their positions and orientations in space.	×
K	SOL.Math.NS.K.1.a	tell how many are in a given set of 20 or fewer objects by counting orally; and	×
K	SOL.Math.NS.K.1.b	read, write, and represent numbers from 0 through 20.	×

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
K	SOL.Math.NS.K.2.a	compare and describe one set as having more, fewer, or the same number of objects as the other set(s); and	X
K	SOL.Math.NS.K.2.b	compare and order sets from least to greatest and greatest to least.	X
K	SOL.Math.NS.K.3.a	count forward orally by ones from 0 to 100;	X
K	SOL.Math.NS.K.3.b	count backward orally by ones when given any number between 1 and 10;	X
K	SOL.Math.NS.K.3.c	identify the number after, without counting, when given any number between 0 and 100 and identify the number before, without counting, when given any number between 1 and 10; and	X
K	SOL.Math.NS.K.3.d	count forward by tens to determine the total number of objects to 100.	X
K	SOL.Math.NS.K.4.a	recognize and describe with fluency part-whole relationships for numbers up to 5; and	X
K	SOL.Math.NS.K.4.b	investigate and describe part-whole relationships for numbers up to 10.	X
K	SOL.Math.NS.K.5	The student will investigate fractions by representing and solving practical problems involving equal sharing with two sharers.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
K	SOL.Math.PFA.K.12	The student will sort and classify objects according to one attribute.	X
K	SOL.Math.PFA.K.13	The student will identify, describe, extend, create, and transfer repeating patterns.	X
K	SOL.Math.PS.K.11.a	collect, organize, and represent data; and	X
K	SOL.Math.PS.K.11.b	read and interpret data in object graphs, picture graphs, and tables.	X
1	SOL.Math.CE.1.6	The student will create and solve single-step story and picture problems using addition and subtraction within 20.	X
1	SOL.Math.CE.1.7.a	recognize and describe with fluency part-whole relationships for numbers up to 10; and	X
1	SOL.Math.CE.1.7.b	demonstrate fluency with addition and subtraction within 10.	X
1	SOL.Math.MG.1.8	The student will determine the value of a collection of like coins (pennies, nickels, or dimes) whose total value is 100 cents or less.	X
1	SOL.Math.MG.1.9.a	tell time to the hour and half-hour, using analog and digital clocks; and	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
1	SOL.Math.MG.1.9.b	read and interpret a calendar.	X
1	SOL.Math.MG.1.10	The student will use nonstandard units to measure and compare length, weight, and volume.	X
1	SOL.Math.MG.1.11.a	identify, trace, describe, and sort plane figures (triangles, squares, rectangles, and circles) according to number of sides, vertices, and angles; and	X
1	SOL.Math.MG.1.11.b	identify and describe representations of circles, squares, rectangles, and triangles in different environments, regardless of orientation, and explain reasoning.	X
1	SOL.Math.NS.1.1.a	count forward orally by ones to 110, starting at any number between 0 and 110;	X
1	SOL.Math.NS.1.1.b	write the numerals 0 to 110 in sequence and out-of-sequence;	X
1	SOL.Math.NS.1.1.c	count backward orally by ones when given any number between 1 and 30; and	X
1	SOL.Math.NS.1.1.d	count forward orally by ones, twos, fives, and tens to determine the total number of objects to 110.	X
1	SOL.Math.NS.1.2.a	group a collection into tens and ones and write the corresponding numeral;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
1	SOL.Math.NS.1.2.b	compare two numbers between 0 and 110 represented pictorially or with concrete objects, using the words greater than, less than or equal to; and	X
1	SOL.Math.NS.1.2.c	order three or fewer sets from least to greatest and greatest to least.	X
1	SOL.Math.NS.1.3	The student, given an ordered set of ten objects and/or pictures, will indicate the ordinal position of each object, first through tenth.	X
1	SOL.Math.NS.1.4.a	represent and solve practical problems involving equal sharing with two or four sharers; and	X
1	SOL.Math.NS.1.4.b	represent and name fractions for halves and fourths, using models.	X
1	SOL.Math.NS.1.5.a	select a reasonable order of magnitude from three given quantities: a one-digit numeral, a two-digit numeral, and a three-digit numeral (e.g., 5, 50, 500); and	X
1	SOL.Math.NS.1.5.b	explain the reasonableness of the choice.	X
1	SOL.Math.PFA.1.13	The student will sort and classify concrete objects according to one or two attributes.	X
1	SOL.Math.PFA.1.14	The student will identify, describe, extend, create, and transfer growing and repeating patterns.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
1	SOL.Math.PFA.1.15	The student will demonstrate an understanding of equality through the use of the equal symbol.	X
1	SOL.Math.PS.1.12.a	collect, organize, and represent various forms of data using tables, picture graphs, and object graphs; and	X
1	SOL.Math.PS.1.12.b	read and interpret data displayed in tables, picture graphs, and object graphs, using the vocabulary more, less, fewer, greater than, less than, and equal to.	X
2	SOL.Math.CE.2.5.a	recognize and use the relationships between addition and subtraction to solve single-step practical problems, with whole numbers to 20; and	X
2	SOL.Math.CE.2.5.b	demonstrate fluency with addition and subtraction within 20.	X
2	SOL.Math.CE.2.6.a	estimate sums and differences;	X
2	SOL.Math.CE.2.6.b	determine sums and differences, using various methods; and	X
2	SOL.Math.CE.2.6.c	create and solve single-step and two-step practical problems involving addition and subtraction.	X
2	SOL.Math.MG.2.7.a	count and compare a collection of pennies, nickels, dimes, and quarters whose total value is $2.00 or less; and	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
2	SOL.Math.MG.2.7.b	use the cent symbol, dollar symbol, and decimal point to write a value of money.	X
2	SOL.Math.MG.2.8.a	length to the nearest inch; and	X
2	SOL.Math.MG.2.8.b	weight to the nearest pound.	X
2	SOL.Math.MG.2.9	The student will tell time and write time to the nearest five minutes, using analog and digital clocks.	X
2	SOL.Math.MG.2.10.a	determine past and future days of the week; and	X
2	SOL.Math.MG.2.10.b	identify specific days and dates on a given calendar.	X
2	SOL.Math.MG.2.11	The student will read temperature to the nearest 10 degrees.	X
2	SOL.Math.MG.2.12.a	draw a line of symmetry in a figure; and	X
2	SOL.Math.MG.2.12.b	identify and create figures with at least one line of symmetry.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
2	SOL.Math.MG.2.13	The student will identify, describe, compare, and contrast plane and solid figures (circles/spheres, squares/cubes, and rectangles/rectangular prisms).	X
2	SOL.Math.NS.2.1.a	read, write, and identify the place and value of each digit in a three-digit numeral, with and without models;	X
2	SOL.Math.NS.2.1.b	identify the number that is 10 more, 10 less, 100 more, and 100 less than a given number up to 999;	X
2	SOL.Math.NS.2.1.c	compare and order whole numbers between 0 and 999; and	X
2	SOL.Math.NS.2.1.d	round two-digit numbers to the nearest ten.	X
2	SOL.Math.NS.2.2.a	count forward by twos, fives, and tens to 120, starting at various multiples of 2, 5, or 10;	X
2	SOL.Math.NS.2.2.b	count backward by tens from 120; and	X
2	SOL.Math.NS.2.2.c	use objects to determine whether a number is even or odd.	X
2	SOL.Math.NS.2.3.a	count and identify the ordinal positions first through twentieth, using an ordered set of objects; and	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
2	SOL.Math.NS.2.3.b	write the ordinal numbers 1st through 20th.	X
2	SOL.Math.NS.2.4.a	name and write fractions represented by a set, region, or length model for halves, fourths, eighths, thirds, and sixths;	X
2	SOL.Math.NS.2.4.b	represent fractional parts with models and with symbols; and	X
2	SOL.Math.NS.2.4.c	compare the unit fractions for halves, fourths, eighths, thirds, and sixths, with models.	X
2	SOL.Math.PFA.2.16	The student will identify, describe, create, extend, and transfer patterns found in objects, pictures, and numbers.	X
2	SOL.Math.PFA.2.17	The student will demonstrate an understanding of equality through the use of the equal symbol and the use of the not equal symbol.	X
2	SOL.Math.PS.2.14	The student will use data from probability experiments to predict outcomes when the experiment is repeated.	X
2	SOL.Math.PS.2.15.a	collect, organize, and represent data in pictographs and bar graphs; and	X
2	SOL.Math.PS.2.15.b	read and interpret data represented in pictographs and bar graphs.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
3	SOL.Math.CE.3.3.a	estimate and determine the sum or difference of two whole numbers; and	X
3	SOL.Math.CE.3.3.b	create and solve single-step and multistep practical problems involving sums or differences of two whole numbers, each 9,999 or less.	X
3	SOL.Math.CE.3.4.a	represent multiplication and division through 10 × 10, using a variety of approaches and models;	X
3	SOL.Math.CE.3.4.b	create and solve single-step practical problems that involve multiplication and division through 10 x 10; and	X
3	SOL.Math.CE.3.4.c	demonstrate fluency with multiplication facts of 0, 1, 2, 5, and 10; and	X
3	SOL.Math.CE.3.4.d	solve single-step practical problems involving multiplication of whole numbers, where one factor is 99 or less and the second factor is 5 or less.	X
3	SOL.Math.CE.3.5	The student will solve practical problems that involve addition and subtraction with proper fractions having like denominators of 12 or less.	X
3	SOL.Math.MG.3.6.a	determine the value of a collection of bills and coins whose total value is $5.00 or less;	X
3	SOL.Math.MG.3.6.b	compare the value of two sets of coins or two sets of coins and bills; and	X

Star Math alignment

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
3	SOL.Math.MG.3.6.c	make change from $5.00 or less.	╳
3	SOL.Math.MG.3.7.a	length to the nearest ½ inch, inch, foot, yard, centimeter, and meter; and	╳
3	SOL.Math.MG.3.7.b	liquid volume in cups, pints, quarts, gallons, and liters.	╳
3	SOL.Math.MG.3.8.a	measure the distance around a polygon in order to determine its perimeter using U.S. Customary and metric units; and	╳
3	SOL.Math.MG.3.8.b	count the number of square units needed to cover a given surface in order to determine its area.	╳
3	SOL.Math.MG.3.9.a	tell time to the nearest minute, using analog and digital clocks;	╳
3	SOL.Math.MG.3.9.b	solve practical problems related to elapsed time in one-hour increments within a 12- hour period; and	╳
3	SOL.Math.MG.3.9.c	identify equivalent periods of time and solve practical problems related to equivalent periods of time.	╳
3	SOL.Math.MG.3.10	The student will read temperature to the nearest degree.	╳

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
3	SOL.Math.MG.3.11	The student will identify and draw representations of points, lines, line segments, rays, and angles.	X
3	SOL.Math.MG.3.12.a	define polygon;	X
3	SOL.Math.MG.3.12.b	identify and name polygons with 10 or fewer sides; and	X
3	SOL.Math.MG.3.12.c	combine and subdivide polygons with three or four sides and name the resulting polygon(s).	X
3	SOL.Math.MG.3.13	The student will identify and describe congruent and noncongruent figures.	X
3	SOL.Math.NS.3.1.a	read, write, and identify the place and value of each digit in a six-digit whole number, with and without models;	X
3	SOL.Math.NS.3.1.b	round whole numbers, 9,999 or less, to the nearest ten, hundred, and thousand; and	X
3	SOL.Math.NS.3.1.c	compare and order whole numbers, each 9,999 or less.	X
3	SOL.Math.NS.3.2.a	name and write fractions and mixed numbers represented by a model;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
3	SOL.Math.NS.3.2.b	represent fractions and mixed numbers with models and symbols; and	X
3	SOL.Math.NS.3.2.c	compare fractions having like and unlike denominators, using words and symbols (>, <, =, or !=), with models.	X
3	SOL.Math.PFA.3.16	The student will identify, describe, create, and extend patterns found in objects, pictures, numbers and tables.	X
3	SOL.Math.PFA.3.17	The student will create equations to represent equivalent mathematical relationships.	X
3	SOL.Math.PS.3.14	The student will investigate and describe the concept of probability as a measurement of chance and list possible outcomes for a single event.	X
3	SOL.Math.PS.3.15.a	collect, organize, and represent data in pictographs or bar graphs; and	X
3	SOL.Math.PS.3.15.b	read and interpret data represented in pictographs and bar graphs.	X
4	SOL.Math.CE.4.4.a	demonstrate fluency with multiplication facts through 12 x 12, and the corresponding division facts;	X
4	SOL.Math.CE.4.4.b	estimate and determine sums, differences, and products of whole numbers;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
4	SOL.Math.CE.4.4.c	estimate and determine quotients of whole numbers, with and without remainders; and	X
4	SOL.Math.CE.4.4.d	create and solve single-step and multistep practical problems involving addition, subtraction, and multiplication, and single-step practical problems involving division with whole numbers.	X
4	SOL.Math.CE.4.5.a	determine common multiples and factors, including least common multiple and greatest common factor;	X
4	SOL.Math.CE.4.5.b	add and subtract fractions and mixed numbers having like and unlike denominators; and	X
4	SOL.Math.CE.4.5.c	solve single-step practical problems involving addition and subtraction with fractions and mixed numbers.	X
4	SOL.Math.CE.4.6.a	add and subtract with decimals; and	X
4	SOL.Math.CE.4.6.b	solve single-step and multistep practical problems involving addition and subtraction with decimals.	X
4	SOL.Math.MG.4.7	The student will solve practical problems that involve determining perimeter and area in U.S. Customary and metric units.	X
4	SOL.Math.MG.4.8.a	estimate and measure length and describe the result in U.S. Customary and metric units;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
4	SOL.Math.MG.4.8.b	estimate and measure weight/mass and describe the result in U.S. Customary and metric units;	X
4	SOL.Math.MG.4.8.c	given the equivalent measure of one unit, identify equivalent measures of length, weight/mass, and liquid volume between units within the U.S. Customary system; and	X
4	SOL.Math.MG.4.8.d	solve practical problems that involve length, weight/mass, and liquid volume in U.S. Customary units.	X
4	SOL.Math.MG.4.9	The student will solve practical problems related to elapsed time in hours and minutes within a 12-hour period.	X
4	SOL.Math.MG.4.10.a	identify and describe points, lines, line segments, rays, and angles, including endpoints and vertices; and	X
4	SOL.Math.MG.4.10.b	identify and describe intersecting, parallel, and perpendicular lines.	X
4	SOL.Math.MG.4.11	The student will identify, describe, compare, and contrast plane and solid figures according to their characteristics (number of angles, vertices, edges, and the number and shape of faces) using concrete models and pictorial representations.	X
4	SOL.Math.MG.4.12	The student will classify quadrilaterals as parallelograms, rectangles, squares, rhombi, and/or trapezoids.	X
4	SOL.Math.NS.4.1.a	read, write, and identify the place and value of each digit in a nine-digit whole number;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
4	SOL.Math.NS.4.1.b	compare and order whole numbers expressed through millions; and	X
4	SOL.Math.NS.4.1.c	round whole numbers expressed through millions to the nearest thousand, ten thousand, and hundred thousand.	X
4	SOL.Math.NS.4.2.a	compare and order fractions and mixed numbers, with and without models;	X
4	SOL.Math.NS.4.2.b	represent equivalent fractions; and	X
4	SOL.Math.NS.4.2.c	identify the division statement that represents a fraction, with models and in context.	X
4	SOL.Math.NS.4.3.a	read, write, represent, and identify decimals expressed through thousandths;	X
4	SOL.Math.NS.4.3.b	round decimals to the nearest whole number;	X
4	SOL.Math.NS.4.3.c	compare and order decimals; and	X
4	SOL.Math.NS.4.3.d	given a model, write the decimal and fraction equivalents.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
4	SOL.Math.PFA.4.15	The student will identify, describe, create, and extend patterns found in objects, pictures, numbers, and tables.	X
4	SOL.Math.PFA.4.16	The student will recognize and demonstrate the meaning of equality in an equation.	X
4	SOL.Math.PS.4.13.a	determine the likelihood of an outcome of a simple event;	X
4	SOL.Math.PS.4.13.b	represent probability as a number between 0 and 1, inclusive; and	X
4	SOL.Math.PS.4.13.c	create a model or practical problem to represent a given probability.	X
4	SOL.Math.PS.4.14.a	collect, organize, and represent data in bar graphs and line graphs;	X
4	SOL.Math.PS.4.14.b	interpret data represented in bar graphs and line graphs; and	X
4	SOL.Math.PS.4.14.c	compare two different representations of the same data (e.g., a set of data displayed on a chart and a bar graph, a chart and a line graph, or a pictograph and a bar graph).	X
5	SOL.Math.CE.5.4	The student will create and solve single-step and multistep practical problems involving addition, subtraction, multiplication, and division of whole numbers.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
5	SOL.Math.CE.5.5.a	estimate and determine the product and quotient of two numbers involving decimals; and	X
5	SOL.Math.CE.5.5.b	create and solve single-step and multistep practical problems involving addition, subtraction, and multiplication of decimals, and create and solve single-step practical problems involving division of decimals.	X
5	SOL.Math.CE.5.6.a	solve single-step and multistep practical problems involving addition and subtraction with fractions and mixed numbers; and	X
5	SOL.Math.CE.5.6.b	solve single-step practical problems involving multiplication of a whole number, limited to 12 or less, and a proper fraction, with models.	X
5	SOL.Math.CE.5.7	The student will simplify whole number numerical expressions using the order of operations.	X
5	SOL.Math.MG.5.8.a	solve practical problems that involve perimeter, area, and volume in standard units of measure; and	X
5	SOL.Math.MG.5.8.b	differentiate among perimeter, area, and volume and identify whether the application of the concept of perimeter, area, or volume is appropriate for a given situation.	X
5	SOL.Math.MG.5.9.a	given the equivalent measure of one unit, identify equivalent measurements within the metric system; and	X
5	SOL.Math.MG.5.9.b	solve practical problems involving length, mass, and liquid volume using metric units.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
5	SOL.Math.MG.5.10	The student will identify and describe the diameter, radius, chord, and circumference of a circle.	X
5	SOL.Math.MG.5.11	The student will solve practical problems related to elapsed time in hours and minutes within a 24-hour period.	X
5	SOL.Math.MG.5.12	The student will classify and measure right, acute, obtuse, and straight angles.	X
5	SOL.Math.MG.5.13.a	classify triangles as right, acute, or obtuse and equilateral, scalene, or isosceles; and	X
5	SOL.Math.MG.5.13.b	investigate the sum of the interior angles in a triangle and determine an unknown angle measure.	X
5	SOL.Math.MG.5.14.a	recognize and apply transformations, such as translation, reflection, and rotation; and	X
5	SOL.Math.MG.5.14.b	investigate and describe the results of combining and subdividing polygons.	X
5	SOL.Math.NS.5.1	The student, given a decimal through thousandths, will round to the nearest whole number, tenth, or hundredth.	X
5	SOL.Math.NS.5.2.a	represent and identify equivalencies among fractions and decimals, with and without models; and	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
5	SOL.Math.NS.5.2.b	compare and order fractions, mixed numbers, and/or decimals in a given set, from least to greatest and greatest to least.	X
5	SOL.Math.NS.5.3.a	identify and describe the characteristics of prime and composite numbers; and	X
5	SOL.Math.NS.5.3.b	identify and describe the characteristics of even and odd numbers.	X
5	SOL.Math.PFA.5.18	The student will identify, describe, create, express, and extend number patterns found in objects, pictures, numbers and tables.	X
5	SOL.Math.PFA.5.19.a	investigate and describe the concept of variable;	X
5	SOL.Math.PFA.5.19.b	write an equation to represent a given mathematical relationship, using a variable;	X
5	SOL.Math.PFA.5.19.c	use an expression with a variable to represent a given verbal expression involving one operation; and	X
5	SOL.Math.PFA.5.19.d	create a problem situation based on a given equation, using a single variable and one operation.	X
5	SOL.Math.PS.5.15	The student will determine the probability of an outcome by constructing a sample space or using the Fundamental (Basic) Counting Principle.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
5	SOL.Math.PS.5.16.a	represent data in line plots and stem-and-leaf plots;	X
5	SOL.Math.PS.5.16.b	interpret data represented in line plots and stem-and-leaf plots; and	X
5	SOL.Math.PS.5.16.c	compare data represented in a line plot with the same data represented in a stem-and-leaf plot.	X
5	SOL.Math.PS.5.17.a	describe mean, median, and mode as measures of center;	X
5	SOL.Math.PS.5.17.b	describe mean as fair share;	X
5	SOL.Math.PS.5.17.c	describe the range of a set of data as a measure of spread; and	X
5	SOL.Math.PS.5.17.d	determine the mean, median, mode, and range of a set of data.	X
6	SOL.Math.CE.6.5.a	multiply and divide fractions and mixed numbers;	X
6	SOL.Math.CE.6.5.b	solve single-step and multistep practical problems involving addition, subtraction, multiplication, and division of fractions and mixed numbers; and	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
6	SOL.Math.CE.6.5.c	solve multistep practical problems involving addition, subtraction, multiplication, and division of decimals.	X
6	SOL.Math.CE.6.6.a	add, subtract, multiply, and divide integers;	X
6	SOL.Math.CE.6.6.b	solve practical problems involving operations with integers; and	X
6	SOL.Math.CE.6.6.c	simplify numerical expressions involving integers.	X
6	SOL.Math.MG.6.7.a	derive ? (pi);	X
6	SOL.Math.MG.6.7.b	solve problems, including practical problems, involving circumference and area of a circle; and	X
6	SOL.Math.MG.6.7.c	solve problems, including practical problems, involving area and perimeter of triangles and rectangles.	X
6	SOL.Math.MG.6.8.a	identify the components of the coordinate plane; and	X
6	SOL.Math.MG.6.8.b	identify the coordinates of a point and graph ordered pairs in a coordinate plane.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
6	SOL.Math.MG.6.9	The student will determine congruence of segments, angles, and polygons.	×
6	SOL.Math.NS.6.1	The student will represent relationships between quantities using ratios, and will use appropriate notations, such as a/b, a to b, and a:b.	×
6	SOL.Math.NS.6.2.a	represent and determine equivalencies among fractions, mixed numbers, decimals, and percents; and	×
6	SOL.Math.NS.6.2.b	compare and order positive rational numbers.	×
6	SOL.Math.NS.6.3.a	identify and represent integers;	×
6	SOL.Math.NS.6.3.b	compare and order integers; and	×
6	SOL.Math.NS.6.3.c	identify and describe absolute value of integers.	×
6	SOL.Math.NS.6.4	The student will recognize and represent patterns with whole number exponents and perfect squares.	×
6	SOL.Math.PFA.6.12.a	represent a proportional relationship between two quantities, including those arising from practical situations;	×

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
6	SOL.Math.PFA.6.12.b	determine the unit rate of a proportional relationship and use it to find a missing value in a ratio table;	X
6	SOL.Math.PFA.6.12.c	determine whether a proportional relationship exists between two quantities; and	X
6	SOL.Math.PFA.6.12.d	make connections between and among representations of a proportional relationship between two quantities using verbal descriptions, ratio tables, and graphs.	X
6	SOL.Math.PFA.6.13	The student will solve one-step linear equations in one variable, including practical problems that require the solution of a one-step linear equation in one variable.	X
6	SOL.Math.PFA.6.14.a	represent a practical situation with a linear inequality in one variable; and	X
6	SOL.Math.PFA.6.14.b	solve one-step linear inequalities in one variable, involving addition or subtraction, and graph the solution on a number line.	X
6	SOL.Math.PS.6.10.a	represent data in a circle graph;	X
6	SOL.Math.PS.6.10.b	make observations and inferences about data represented in a circle graph; and	X
6	SOL.Math.PS.6.10.c	compare circle graphs with the same data represented in bar graphs, pictographs, and line plots.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
6	SOL.Math.PS.6.11.a	represent the mean of a data set graphically as the balance point; and	X
6	SOL.Math.PS.6.11.b	determine the effect on measures of center when a single value of a data set is added, removed, or changed.	X
7	SOL.Math.CE.7.2	The student will solve practical problems involving operations with rational numbers.	X
7	SOL.Math.CE.7.3	The student will solve single-step and multistep practical problems, using proportional reasoning.	X
7	SOL.Math.MG.7.4.a	describe and determine the volume and surface area of rectangular prisms and cylinders; and	X
7	SOL.Math.MG.7.4.b	solve problems, including practical problems, involving the volume and surface area of rectangular prisms and cylinders.	X
7	SOL.Math.MG.7.5	The student will solve problems, including practical problems, involving the relationship between corresponding sides and corresponding angles of similar quadrilaterals and triangles.	X
7	SOL.Math.MG.7.6.a	compare and contrast quadrilaterals based on their properties; and	X
7	SOL.Math.MG.7.6.b	determine unknown side lengths or angle measures of quadrilaterals.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
7	SOL.Math.MG.7.7	The student will apply translations and reflections of right triangles or rectangles in the coordinate plane.	X
7	SOL.Math.NS.7.1.a	investigate and describe the concept of negative exponents for powers of ten;	X
7	SOL.Math.NS.7.1.b	compare and order numbers greater than zero written in scientific notation;	X
7	SOL.Math.NS.7.1.c	compare and order rational numbers;	X
7	SOL.Math.NS.7.1.d	determine square roots of perfect squares; and	X
7	SOL.Math.NS.7.1.e	identify and describe absolute value of rational numbers.	X
7	SOL.Math.PFA.7.10.a	determine the slope, m, as rate of change in a proportional relationship between two quantities and write an equation in the form y = mx to represent the relationship;	X
7	SOL.Math.PFA.7.10.b	graph a line representing a proportional relationship between two quantities given the slope and an ordered pair, or given the equation in y = mx form where m represents the slope as rate of change;	X
7	SOL.Math.PFA.7.10.c	determine the y-intercept, b, in an additive relationship between two quantities and write an equation in the form y = x + b to represent the relationship;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
7	SOL.Math.PFA.7.10.d	graph a line representing an additive relationship between two quantities given the y-intercept and an ordered pair, or given the equation in the form y = x + b, where b represents the y-intercept; and	X
7	SOL.Math.PFA.7.10.e	make connections between and among representations of a proportional or additive relationship between two quantities using verbal descriptions, tables, equations, and graphs.	X
7	SOL.Math.PFA.7.11	The student will evaluate algebraic expressions for given replacement values of the variables.	X
7	SOL.Math.PFA.7.12	The student will solve two-step linear equations in one variable, including practical problems that require the solution of a two-step linear equation in one variable.	X
7	SOL.Math.PFA.7.13	The student will solve one- and two-step linear inequalities in one variable, including practical problems, involving addition, subtraction, multiplication, and division, and graph the solution on a number line.	X
7	SOL.Math.PS.7.8.a	determine the theoretical and experimental probabilities of an event; and	X
7	SOL.Math.PS.7.8.b	investigate and describe the difference between the experimental probability and theoretical probability of an event.	X
7	SOL.Math.PS.7.9.a	represent data in a histogram;	X
7	SOL.Math.PS.7.9.b	make observations and inferences about data represented in a histogram; and	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
7	SOL.Math.PS.7.9.c	compare histograms with the same data represented in stem-and-leaf plots, line plots, and circle graphs.	×
8	SOL.Math.CE.8.4	The student will solve practical problems involving consumer applications.	×
8	SOL.Math.MG.8.5	The student will use the relationships among pairs of angles that are vertical angles, adjacent angles, supplementary angles, and complementary angles to determine the measure of unknown angles.	×
8	SOL.Math.MG.8.6.a	solve problems, including practical problems, involving volume and surface area of cones and square-based pyramids; and	×
8	SOL.Math.MG.8.6.b	describe how changing one measured attribute of a rectangular prism affects the volume and surface area.	×
8	SOL.Math.MG.8.7.a	given a polygon, apply transformations, to include translations, reflections, and dilations, in the coordinate plane; and	×
8	SOL.Math.MG.8.7.b	identify practical applications of transformations.	×
8	SOL.Math.MG.8.8	The student will construct a three-dimensional model, given the top or bottom, side, and front views.	×
8	SOL.Math.MG.8.9.a	verify the Pythagorean Theorem; and	×

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
8	SOL.Math.MG.8.9.b	apply the Pythagorean Theorem.	X
8	SOL.Math.MG.8.10	The student will solve area and perimeter problems, including practical problems, involving composite plane figures.	X
8	SOL.Math.NS.8.1	The student will compare and order real numbers.	X
8	SOL.Math.NS.8.2	The student will describe the relationships between the subsets of the real number system.	X
8	SOL.Math.NS.8.3.a	estimate and determine the two consecutive integers between which a square root lies; and	X
8	SOL.Math.NS.8.3.b	determine both the positive and negative square roots of a given perfect square.	X
8	SOL.Math.PFA.8.14.a	evaluate an algebraic expression for given replacement values of the variables; and	X
8	SOL.Math.PFA.8.14.b	simplify algebraic expressions in one variable.	X
8	SOL.Math.PFA.8.15.a	determine whether a given relation is a function; and	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
8	SOL.Math.PFA.8.15.b	determine the domain and range of a function.	X
8	SOL.Math.PFA.8.16.a	recognize and describe the graph of a linear function with a slope that is positive, negative, or zero;	X
8	SOL.Math.PFA.8.16.b	identify the slope and y-intercept of a linear function, given a table of values, a graph, or an equation in y = mx + b form;	X
8	SOL.Math.PFA.8.16.c	determine the independent and dependent variable, given a practical situation modeled by a linear function;	X
8	SOL.Math.PFA.8.16.d	graph a linear function given the equation in y = mx + b form; and	X
8	SOL.Math.PFA.8.16.e	make connections between and among representations of a linear function using verbal descriptions, tables, equations, and graphs.	X
8	SOL.Math.PFA.8.17	The student will solve multistep linear equations in one variable with the variable on one or both sides of the equation, including practical problems that require the solution of a multistep linear equation in one variable.	X
8	SOL.Math.PFA.8.18	The student will solve multistep linear inequalities in one variable with the variable on one or both sides of the inequality symbol, including practical problems, and graph the solution on a number line.	X
8	SOL.Math.PS.8.11.a	compare and contrast the probability of independent and dependent events; and	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
8	SOL.Math.PS.8.11.b	determine probabilities for independent and dependent events.	×
8	SOL.Math.PS.8.12.a	represent numerical data in boxplots;	×
8	SOL.Math.PS.8.12.b	make observations and inferences about data represented in boxplots; and	×
8	SOL.Math.PS.8.12.c	compare and analyze two data sets using boxplots.	×
8	SOL.Math.PS.8.13.a	represent data in scatterplots;	×
8	SOL.Math.PS.8.13.b	make observations about data represented in scatterplots; and	×
8	SOL.Math.PS.8.13.c	use a drawing to estimate the line of best fit for data represented in a scatterplot.	×
A1	SOL.Math.EI.A.4.a	multistep linear equations in one variable algebraically;	×
A1	SOL.Math.EI.A.4.b	quadratic equations in one variable algebraically;	×

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
A1	SOL.Math.EI.A.4.c	literal equations for a specified variable;	X
A1	SOL.Math.EI.A.4.d	systems of two linear equations in two variables algebraically and graphically; and	X
A1	SOL.Math.EI.A.4.e	practical problems involving equations and systems of equations.	X
A1	SOL.Math.EI.A.5.a	solve multistep linear inequalities in one variable algebraically and represent the solution graphically;	X
A1	SOL.Math.EI.A.5.b	represent the solution of linear inequalities in two variables graphically;	X
A1	SOL.Math.EI.A.5.c	solve practical problems involving inequalities; and	X
A1	SOL.Math.EI.A.5.d	represent the solution to a system of inequalities graphically.	X
A1	SOL.Math.EI.A.6.a	determine the slope of a line when given an equation of the line, the graph of the line, or two points on the line;	X
A1	SOL.Math.EI.A.6.b	write the equation of a line when given the graph of the line, two points on the line, or the slope and a point on the line; and	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
A1	SOL.Math.EI.A.6.c	graph linear equations in two variables.	X
A1	SOL.Math.EO.A.1.a	represent verbal quantitative situations algebraically; and	X
A1	SOL.Math.EO.A.1.b	evaluate algebraic expressions for given replacement values of the variables.	X
A1	SOL.Math.EO.A.2.a	applying the laws of exponents to perform operations on expressions;	X
A1	SOL.Math.EO.A.2.b	adding, subtracting, multiplying, and dividing polynomials; and	X
A1	SOL.Math.EO.A.2.c	factoring completely first- and second-degree binomials and trinomials in one variable.	X
A1	SOL.Math.EO.A.3.a	square roots of whole numbers and monomial algebraic expressions;	X
A1	SOL.Math.EO.A.3.b	cube roots of integers; and	X
A1	SOL.Math.EO.A.3.c	numerical expressions containing square or cube roots.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
A1	SOL.Math.F.A.7.a	determining whether a relation is a function;	X
A1	SOL.Math.F.A.7.b	domain and range;	X
A1	SOL.Math.F.A.7.c	zeros;	X
A1	SOL.Math.F.A.7.d	intercepts;	X
A1	SOL.Math.F.A.7.e	values of a function for elements in its domain; and	X
A1	SOL.Math.F.A.7.f	connections between and among multiple representations of functions using verbal descriptions, tables, equations, and graphs.	X
A1	SOL.Math.S.A.8	The student, given a data set or practical situation, will analyze a relation to determine whether a direct or inverse variation exists, and represent a direct variation algebraically and graphically and an inverse variation algebraically.	X
A1	SOL.Math.S.A.9	The student will collect and analyze data, determine the equation of the curve of best fit in order to make predictions, and solve practical problems, using mathematical models of linear and quadratic functions.	X
GM	SOL.Math.PC.G.9	The student will verify and use properties of quadrilaterals to solve problems, including practical problems.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
GM	SOL.Math.PC.G.10.a	sum of the interior and/or exterior angles;	X
GM	SOL.Math.PC.G.10.b	measure of an interior and/or exterior angle; and	X
GM	SOL.Math.PC.G.10.c	number of sides of a regular polygon.	X
GM	SOL.Math.PC.G.11.a	angle measures formed by intersecting chords, secants, and/or tangents;	X
GM	SOL.Math.PC.G.11.b	lengths of segments formed by intersecting chords, secants, and/or tangents;	X
GM	SOL.Math.PC.G.11.c	arc length; and	X
GM	SOL.Math.PC.G.11.d	area of a sector.	X
GM	SOL.Math.PC.G.12	The student will solve problems involving equations of circles.	X
GM	SOL.Math.RLT.G.1.a	identifying the converse, inverse, and contrapositive of a conditional statement;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
GM	SOL.Math.RLT.G.1.b	translating a short verbal argument into symbolic form; and	X
GM	SOL.Math.RLT.G.1.c	determining the validity of a logical argument.	X
GM	SOL.Math.RLT.G.2.a	prove two or more lines are parallel; and	X
GM	SOL.Math.RLT.G.2.b	solve problems, including practical problems, involving angles formed when parallel lines are intersected by a transversal.	X
GM	SOL.Math.RLT.G.3.a	investigating and using formulas for determining distance, midpoint, and slope;	X
GM	SOL.Math.RLT.G.3.b	applying slope to verify and determine whether lines are parallel or perpendicular;	X
GM	SOL.Math.RLT.G.3.c	investigating symmetry and determining whether a figure is symmetric with respect to a line or a point; and	X
GM	SOL.Math.RLT.G.3.d	determining whether a figure has been translated, reflected, rotated, or dilated, using coordinate methods.	X
GM	SOL.Math.RLT.G.4.a	a line segment congruent to a given line segment;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
GM	SOL.Math.RLT.G.4.b	the perpendicular bisector of a line segment;	X
GM	SOL.Math.RLT.G.4.c	a perpendicular to a given line from a point not on the line;	X
GM	SOL.Math.RLT.G.4.d	a perpendicular to a given line at a given point on the line;	X
GM	SOL.Math.RLT.G.4.e	the bisector of a given angle,	X
GM	SOL.Math.RLT.G.4.f	an angle congruent to a given angle;	X
GM	SOL.Math.RLT.G.4.g	a line parallel to a given line through a point not on the line; and	X
GM	SOL.Math.RLT.G.4.h	an equilateral triangle, a square, and a regular hexagon inscribed in a circle.	X
GM	SOL.Math.T.G.5.a	ordering the sides by length, given angle measures;	X
GM	SOL.Math.T.G.5.b	ordering the angles by degree measure, given side lengths;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
GM	SOL.Math.T.G.5.c	determining whether a triangle exists; and	X
GM	SOL.Math.T.G.5.d	determining the range in which the length of the third side must lie.	X
GM	SOL.Math.T.G.6	The student, given information in the form of a figure or statement, will prove two triangles are congruent.	X
GM	SOL.Math.T.G.7	The student, given information in the form of a figure or statement, will prove two triangles are similar.	X
GM	SOL.Math.T.G.8.a	the Pythagorean Theorem and its converse;	X
GM	SOL.Math.T.G.8.b	properties of special right triangles; and	X
GM	SOL.Math.T.G.8.c	trigonometric ratios.	X
GM	SOL.Math.TDF.G.13	The student will use surface area and volume of three-dimensional objects to solve practical problems.	X
GM	SOL.Math.TDF.G.14.a	comparing ratios between lengths, perimeters, areas, and volumes of similar figures;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
GM	SOL.Math.TDF.G.14.b	determining how changes in one or more dimensions of a figure affect area and/or volume of the figure;	X
GM	SOL.Math.TDF.G.14.c	determining how changes in area and/or volume of a figure affect one or more dimensions of the figure; and	X
GM	SOL.Math.TDF.G.14.d	solving problems, including practical problems, about similar geometric figures.	X
A2	SOL.Math.EI.AII.3.a	absolute value linear equations and inequalities;	X
A2	SOL.Math.EI.AII.3.b	quadratic equations over the set of complex numbers;	X
A2	SOL.Math.EI.AII.3.c	equations containing rational algebraic expressions; and	X
A2	SOL.Math.EI.AII.3.d	equations containing radical expressions.	X
A2	SOL.Math.EI.AII.4	The student will solve systems of linear-quadratic and quadratic-quadratic equations, algebraically and graphically.	X
A2	SOL.Math.EO.AII.1.a	add, subtract, multiply, divide, and simplify rational algebraic expressions;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
A2	SOL.Math.EO.AII.1.b	add, subtract, multiply, divide, and simplify radical expressions containing rational numbers and variables, and expressions containing rational exponents; and	X
A2	SOL.Math.EO.AII.1.c	factor polynomials completely in one or two variables.	X
A2	SOL.Math.EO.AII.2	The student will perform operations on complex numbers and express the results in simplest form using patterns of the powers of i.	X
A2	SOL.Math.F.AII.5	The student will investigate and apply the properties of arithmetic and geometric sequences and series to solve practical problems, including writing the first n terms, determining then nth term, and evaluating summation formulas. Notation will include ? and an.	X
A2	SOL.Math.F.AII.6.a	recognize the general shape of function families; and	X
A2	SOL.Math.F.AII.6.b	use knowledge of transformations to convert between equations and the corresponding graphs of functions.	X
A2	SOL.Math.F.AII.7.a	domain, range, and continuity;	X
A2	SOL.Math.F.AII.7.b	intervals in which a function is increasing or decreasing;	X
A2	SOL.Math.F.AII.7.c	extrema;	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
A2	SOL.Math.F.AII.7.d	zeros;	X
A2	SOL.Math.F.AII.7.e	intercepts;	X
A2	SOL.Math.F.AII.7.f	values of a function for elements in its domain;	X
A2	SOL.Math.F.AII.7.g	connections between and among multiple representations of functions using verbal descriptions, tables, equations, and graphs;	X
A2	SOL.Math.F.AII.7.h	end behavior;	X
A2	SOL.Math.F.AII.7.i	vertical and horizontal asymptotes;	X
A2	SOL.Math.F.AII.7.j	inverse of a function; and	X
A2	SOL.Math.F.AII.7.k	composition of functions algebraically and graphically.	X
A2	SOL.Math.F.AII.8	The student will investigate and describe the relationships among solutions of an equation, zeros of a function, x-intercepts of a graph, and factors of a polynomial expression.	X

Renaissance Star Math Aligns to the Mathematics Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Math Alignment
A2	SOL.Math.S.AII.9	The student will collect and analyze data, determine the equation of the curve of best fit in order to make predictions, and solve practical problems, using mathematical models of quadratic and exponential functions.	╳
A2	SOL.Math.S.AII.10	The student will represent and solve problems, including practical problems, involving inverse variation, joint variation, and a combination of direct and inverse variations.	╳
A2	SOL.Math.S.AII.11.a	identify and describe properties of a normal distribution;	╳
A2	SOL.Math.S.AII.11.b	interpret and compare z-scores for normally distributed data; and	╳
A2	SOL.Math.S.AII.11.c	apply properties of normal distributions to determine probabilities associated with areas under the standard normal curve.	╳
A2	SOL.Math.S.AII.12	The student will compute and distinguish between permutations and combinations.	╳

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
K	SOL.English.K.3.a	Begin to discriminate between spoken sentences, words, and syllables.	×
K	SOL.English.K.3.b	Identify and produce words that rhyme.	×
K	SOL.English.K.3.c	Blend and segment multisyllabic words at the syllable level.	×
K	SOL.English.K.3.d	Blend and segment one-syllable words into phonemes including onset and rime.	×
K	SOL.English.K.3.e	Identify words according to shared beginning and/or ending sounds.	×
K	SOL.English.K.3.f	Blend sounds to make one-syllable words.	×
K	SOL.English.K.3.g	Segment one-syllable words into individual phonemes.	×
K	SOL.English.K.4.a	Hold print materials in the correct position.	×
K	SOL.English.K.4.b	Identify the front cover, back cover, and title page of a book.	×
K	SOL.English.K.4.c	Distinguish between print and pictures.	×
K	SOL.English.K.4.d	Follow words from left to right and from top to bottom on a printed page.	×
K	SOL.English.K.4.e	Match voice with print.	×
K	SOL.English.K.5.a	Identify common signs and logos.	×

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
K	SOL.English.K.5.b	Explain that printed materials provide information.	X
K	SOL.English.K.5.c	Read and explain own writing and drawings.	X
K	SOL.English.K.5.d	Read his/her name and commonly used high-frequency words.	X
K	SOL.English.K.6.a	Identify and name the capital and lowercase letters of the alphabet.	X
K	SOL.English.K.6.b	Match consonant, short vowel, and initial consonant digraph sounds to appropriate letters.	X
K	SOL.English.K.6.c	Demonstrate a speech-to-print match through accurate finger-point reading in familiar text that includes words with more than one syllable.	X
K	SOL.English.K.6.d	Identify initial consonant sounds in one-syllable words.	X
K	SOL.English.K.6.e	Identify final consonant sounds in one-syllable words.	X
K	SOL.English.K.7.a	Discuss meanings of words.	X
K	SOL.English.K.7.b	Increase vocabulary by listening to a variety of texts read aloud.	X
K	SOL.English.K.7.c	Use vocabulary from other content areas.	X
K	SOL.English.K.7.d	Ask about words not understood.	X
K	SOL.English.K.7.e	Use number words.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
K	SOL.English.K.7.f	Use nouns to identify and name people, places, and things.	X
K	SOL.English.K.7.g	Use adjectives to describe location, size, color, and shape.	X
K	SOL.English.K.7.h	Use verbs to identify actions.	X
K	SOL.English.K.8	The student will demonstrate comprehension of fictional texts.	X
K	SOL.English.K.8.a	Identify the role of an author and an illustrator.	X
K	SOL.English.K.8.b	Relate previous experiences to what is read.	X
K	SOL.English.K.8.c	Use pictures to make predictions.	X
K	SOL.English.K.8.d	Ask and answer questions about what is read.	X
K	SOL.English.K.8.e	Use story elements of characters, settings, and events to retell stories sequentially using beginning, middle, and end.	X
K	SOL.English.K.9	The student will demonstrate comprehension of nonfiction texts.	X
K	SOL.English.K.9.a	Use pictures to identify topic and make predictions.	X
K	SOL.English.K.9.b	Identify text features specific to the topic, such as titles, headings, and pictures.	X
K	SOL.English.K.9.c	Ask and answer questions about what is read.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
1	SOL.English.1.3.a	Create rhyming words.	X
1	SOL.English.1.3.b	Count phonemes (sounds) in one-syllable words.	X
1	SOL.English.1.3.c	Blend sounds to make one-syllable words.	X
1	SOL.English.1.3.d	Segment one-syllable words into individual phonemes.	X
1	SOL.English.1.3.e	Add or delete phonemes to make new words.	X
1	SOL.English.1.3.f	Blend and segment multisyllabic words at the syllable level.	X
1	SOL.English.1.4.a	Read from left to right and from top to bottom.	X
1	SOL.English.1.4.b	Match spoken words with print.	X
1	SOL.English.1.4.c	Identify letters, words, sentences, and ending punctuation.	X
1	SOL.English.1.5.a	Use initial and final consonants to decode and spell one-syllable words.	X
1	SOL.English.1.5.b	Use two-letter consonant blends to decode and spell one-syllable words.	X
1	SOL.English.1.5.c	Use consonant digraphs to decode and spell one-syllable words.	X
1	SOL.English.1.5.d	Use short vowel sounds to decode and spell one-syllable words.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
1	SOL.English.1.5.e	Blend initial, medial, and final sounds to recognize and read words.	X
1	SOL.English.1.5.f	Use word patterns to decode unfamiliar words.	X
1	SOL.English.1.5.g	Read and spell simple two-syllable compound words.	X
1	SOL.English.1.5.h	Read and spell commonly used sight words.	X
1	SOL.English.1.6	The student will use semantic clues and syntax for support when reading.	X
1	SOL.English.1.6.a	Use words, phrases, and sentences.	X
1	SOL.English.1.6.b	Use titles and pictures.	X
1	SOL.English.1.6.c	Use information in the story to read words.	X
1	SOL.English.1.6.d	Use knowledge of sentence structure.	X
1	SOL.English.1.6.e	Reread and self-correct.	X
1	SOL.English.1.7.a	Discuss meanings of words in context.	X
1	SOL.English.1.7.b	Develop vocabulary by listening to and reading a variety of texts.	X
1	SOL.English.1.7.c	Ask for the meaning of unknown words and make connections to familiar words.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
1	SOL.English.1.7.d	Use text clues such as words or pictures to discern meanings of unknown words.	X
1	SOL.English.1.7.e	Use vocabulary from other content areas.	X
1	SOL.English.1.7.f	Use singular and plural nouns.	X
1	SOL.English.1.7.g	Use adjectives to describe nouns.	X
1	SOL.English.1.7.h	Use verbs to identify actions.	X
1	SOL.English.1.8.a	Use knowledge of alphabetical order by first letter.	X
1	SOL.English.1.8.b	Use a picture dictionary to find meanings of unfamiliar words.	X
1	SOL.English.1.9	The student will read and demonstrate comprehension of a variety of fictional texts.	X
1	SOL.English.1.9.a	Preview the selection.	X
1	SOL.English.1.9.b	Set a purpose for reading.	X
1	SOL.English.1.9.c	Relate previous experiences to what is read.	X
1	SOL.English.1.9.d	Make and confirm predictions.	X
1	SOL.English.1.9.e	Ask and answer who, what, when, where, why, and how questions about what is read.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
1	SOL.English.1.9.f	Identify characters, setting, and important events.	X
1	SOL.English.1.9.g	Retell stories and events, using beginning, middle, and end in a sequential order.	X
1	SOL.English.1.9.h	Identify theme.	X
1	SOL.English.1.9.i	Read and reread familiar stories and poems with fluency, accuracy, and meaningful expression.	X
1	SOL.English.1.10	The student will read and demonstrate comprehension of a variety of nonfiction texts.	X
1	SOL.English.1.10.a	Preview the selection.	X
1	SOL.English.1.10.b	Use prior and background knowledge as context for new learning.	X
1	SOL.English.1.10.c	Set a purpose for reading.	X
1	SOL.English.1.10.d	Identify text features such as pictures, headings, charts, and captions.	X
1	SOL.English.1.10.e	Make and confirm predictions.	X
1	SOL.English.1.10.f	Ask and answer who, what, where, when, why, and how questions about what is read.	X
1	SOL.English.1.10.g	Identify the main idea.	X
1	SOL.English.1.10.h	Read and reread familiar texts with fluency, accuracy, and meaningful expression.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
2	SOL.English.2.3.a	Count phonemes within one-syllable words.	X
2	SOL.English.2.3.b	Blend sounds to make one-syllable words.	X
2	SOL.English.2.3.c	Segment one-syllable words into phonemes.	X
2	SOL.English.2.3.d	Add or delete phonemes to make words.	X
2	SOL.English.2.3.e	Blend and segment multisyllabic words at the syllable level.	X
2	SOL.English.2.4.a	Use knowledge of consonants, consonant blends, and consonant digraphs to decode and spell words.	X
2	SOL.English.2.4.b	Use knowledge of short, long, and r-controlled vowel patterns to decode and spell words.	X
2	SOL.English.2.4.c	Decode regular multisyllabic words.	X
2	SOL.English.2.4.d	Apply decoding strategies to confirm or correct while reading.	X
2	SOL.English.2.5.a	Use information and context clues in the story to read words.	X
2	SOL.English.2.5.b	Use knowledge of sentence structure to determine the meaning of unknown words.	X
2	SOL.English.2.6.a	Use knowledge of homophones.	X
2	SOL.English.2.6.b	Use knowledge of prefixes and suffixes.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
2	SOL.English.2.6.c	Use knowledge of antonyms and synonyms.	X
2	SOL.English.2.6.d	Discuss meanings of words and develop vocabulary by listening to and reading a variety of texts.	X
2	SOL.English.2.6.e	Use word-reference materials including dictionaries, glossaries and indices.	X
2	SOL.English.2.6.f	Use vocabulary from other content areas.	X
2	SOL.English.2.7	The student will read and demonstrate comprehension of fictional texts.	X
2	SOL.English.2.7.a	Make and confirm predictions.	X
2	SOL.English.2.7.b	Connect previous experiences to new texts.	X
2	SOL.English.2.7.c	Ask and answer questions using the text for support.	X
2	SOL.English.2.7.d	Describe characters, setting, and plot events in fiction and poetry.	X
2	SOL.English.2.7.e	Identify the conflict and resolution.	X
2	SOL.English.2.7.f	Identify the theme.	X
2	SOL.English.2.7.g	Summarize stories and events with beginning, middle, and end in the correct sequence.	X
2	SOL.English.2.7.h	Draw conclusions based on the text.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
2	SOL.English.2.7.i	Read and reread familiar stories and poems with fluency, accuracy, and meaningful expression.	X
2	SOL.English.2.8	The student will read and demonstrate comprehension of nonfiction texts.	X
2	SOL.English.2.8.a	Preview the selection using text features including table of contents, headings, pictures, captions, and maps.	X
2	SOL.English.2.8.b	Make and confirm predictions.	X
2	SOL.English.2.8.c	Use prior and background knowledge as context for new learning.	X
2	SOL.English.2.8.d	Set purpose for reading.	X
2	SOL.English.2.8.e	Ask and answer questions using the text as support.	X
2	SOL.English.2.8.f	Identify the main idea.	X
2	SOL.English.2.8.g	Draw conclusions based on the text.	X
2	SOL.English.2.8.h	Read and reread familiar texts with fluency, accuracy, and meaningful expression.	X
3	SOL.English.3.3.a	Use knowledge of regular and irregular vowel patterns.	X
3	SOL.English.3.3.b	Decode regular multisyllabic words.	X
3	SOL.English.3.4.a	Use knowledge of homophones.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
3	SOL.English.3.4.b	Use knowledge of roots, affixes, synonyms, and antonyms to determine the meaning of new words.	X
3	SOL.English.3.4.c	Apply meaning clues, language structure, and phonetic strategies to determine the meaning of new words.	X
3	SOL.English.3.4.d	Use context to clarify meaning of unfamiliar words.	X
3	SOL.English.3.4.e	Discuss meanings of words and develop vocabulary by listening to and reading a variety of texts.	X
3	SOL.English.3.4.f	Use vocabulary from other content areas.	X
3	SOL.English.3.4.g	Use word-reference resources including the glossary, dictionary, and thesaurus.	X
3	SOL.English.3.5	The student will read and demonstrate comprehension of fictional texts, literary nonfiction, and poetry.	X
3	SOL.English.3.5.a	Set a purpose for reading.	X
3	SOL.English.3.5.b	Make connections between reading selections.	X
3	SOL.English.3.5.c	Make, confirm, and revise predictions.	X
3	SOL.English.3.5.d	Compare and contrast settings, characters, and plot events.	X
3	SOL.English.3.5.e	Summarize plot events.	X
3	SOL.English.3.5.f	Identify the narrator of a story.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
3	SOL.English.3.5.g	Ask and answer questions about what is read.	X
3	SOL.English.3.5.h	Draw conclusions using the text for support.	X
3	SOL.English.3.5.i	Identify the conflict and resolution.	X
3	SOL.English.3.5.j	Identify the theme.	X
3	SOL.English.3.5.k	Use reading strategies to monitor comprehension throughout the reading process.	X
3	SOL.English.3.5.l	Differentiate between fiction and nonfiction.	X
3	SOL.English.3.5.m	Read with fluency, accuracy, and meaningful expression.	X
3	SOL.English.3.6	The student will read and demonstrate comprehension of nonfiction texts.	X
3	SOL.English.3.6.a	Identify the author's purpose.	X
3	SOL.English.3.6.b	Use prior and background knowledge as context for new learning.	X
3	SOL.English.3.6.c	Preview and use text features including table of contents, headings, pictures, captions, maps, indices, and charts.	X
3	SOL.English.3.6.d	Ask and answer questions about what is read using the text for support.	X
3	SOL.English.3.6.e	Draw conclusions using the text for support.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
3	SOL.English.3.6.f	Summarize information found in nonfiction texts.	X
3	SOL.English.3.6.g	Identify the main idea.	X
3	SOL.English.3.6.h	Identify supporting details.	X
3	SOL.English.3.6.i	Use reading strategies to monitor comprehension throughout the reading process.	X
3	SOL.English.3.6.j	Read with fluency, accuracy, and meaningful expression.	X
4	SOL.English.4.4.a	Use context to clarify meanings of unfamiliar words.	X
4	SOL.English.4.4.b	Use knowledge of roots, affixes, synonyms, antonyms, and homophones to determine the meaning of new words.	X
4	SOL.English.4.4.c	Use word-reference materials.	X
4	SOL.English.4.4.d	Use vocabulary from other content areas.	X
4	SOL.English.4.4.e	Develop and use general and specialized vocabulary through speaking, listening, reading, and writing.	X
4	SOL.English.4.5	The student will read and demonstrate comprehension of fictional texts, literary nonfiction texts, and poetry.	X
4	SOL.English.4.5.a	Describe how the choice of language, setting, and characters contributes to the development of plot.	X
4	SOL.English.4.5.b	Identify the theme(s).	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
4	SOL.English.4.5.c	Summarize events in the plot.	X
4	SOL.English.4.5.d	Identify genres.	X
4	SOL.English.4.5.e	Identify the narrator of a story and the speaker of a poem.	X
4	SOL.English.4.5.f	Identify the conflict and resolution.	X
4	SOL.English.4.5.g	Identify sensory words.	X
4	SOL.English.4.5.h	Draw conclusions/make inferences about text using the text as support.	X
4	SOL.English.4.5.i	Compare/contrast details in literary and informational nonfiction texts.	X
4	SOL.English.4.5.j	Identify cause and effect relationships.	X
4	SOL.English.4.5.k	Use reading strategies throughout the reading process to monitor comprehension.	X
4	SOL.English.4.5.l	Read with fluency, accuracy, and meaningful expression.	X
4	SOL.English.4.6	The student will read and demonstrate comprehension of nonfiction texts.	X
4	SOL.English.4.6.a	Use text features such as type, headings, and graphics, to predict and categorize information.	X
4	SOL.English.4.6.b	Explain the author's purpose.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
4	SOL.English.4.6.c	Identify the main idea.	X
4	SOL.English.4.6.d	Summarize supporting details.	X
4	SOL.English.4.6.e	Draw conclusions and make inferences using textual information as support.	X
4	SOL.English.4.6.f	Distinguish between cause and effect.	X
4	SOL.English.4.6.g	Distinguish between fact and opinion.	X
4	SOL.English.4.6.h	Use reading strategies throughout the reading process to monitor comprehension.	X
4	SOL.English.4.6.i	Read with fluency, accuracy, and meaningful expression.	X
5	SOL.English.5.4.a	Use context to clarify meaning of unfamiliar words and phrases.	X
5	SOL.English.5.4.b	Use context and sentence structure to determine meanings and differentiate among multiple meanings of words.	X
5	SOL.English.5.4.c	Use knowledge of roots, affixes, synonyms, antonyms, and homophones to determine the meaning of new words.	X
5	SOL.English.5.4.d	Identify an author’s use of figurative language.	X
5	SOL.English.5.4.e	Use word-reference materials.	X
5	SOL.English.5.4.f	Develop and use general and specialized content area vocabulary through speaking, listening, reading, and writing.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
5	SOL.English.5.5	The student will read and demonstrate comprehension of fictional texts, literary nonfiction, and poetry.	X
5	SOL.English.5.5.a	Summarize plot events using details from text.	X
5	SOL.English.5.5.b	Discuss the impact of setting on plot development.	X
5	SOL.English.5.5.c	Describe character development.	X
5	SOL.English.5.5.d	Identify theme(s).	X
5	SOL.English.5.5.e	Explain the resolution of conflict(s).	X
5	SOL.English.5.5.f	Identify genres.	X
5	SOL.English.5.5.g	Differentiate between first and third person point-of-view.	X
5	SOL.English.5.5.h	Differentiate between free verse and rhymed poetry.	X
5	SOL.English.5.5.i	Explain how an author’s choice of vocabulary contributes to the author’s style.	X
5	SOL.English.5.5.j	Draw conclusions and make inferences with support from the text.	X
5	SOL.English.5.5.k	Identify cause and effect relationships.	X
5	SOL.English.5.5.l	Compare/contrast details in literary and informational nonfiction texts.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
5	SOL.English.5.5.m	Use reading strategies throughout the reading process to monitor comprehension.	X
5	SOL.English.5.6	The student will read and demonstrate comprehension of nonfiction texts.	X
5	SOL.English.5.6.a	Use text features such as type, headings, and graphics, to predict and categorize information.	X
5	SOL.English.5.6.b	Skim materials to develop a general overview of content and to locate specific information.	X
5	SOL.English.5.6.c	Identify the main idea.	X
5	SOL.English.5.6.d	Summarize supporting details.	X
5	SOL.English.5.6.e	Identify organizational pattern(s).	X
5	SOL.English.5.6.f	Identify transitional words and phrases that signal an author’s organizational pattern.	X
5	SOL.English.5.6.g	Locate information from the text to support opinions, inferences, and conclusions.	X
5	SOL.English.5.6.h	Identify cause and effect relationships.	X
5	SOL.English.5.6.i	Differentiate between fact and opinion.	X
5	SOL.English.5.6.j	Compare and contrast details and ideas within and between texts.	X
5	SOL.English.5.6.k	Use reading strategies throughout the reading process to monitor comprehension.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
6	SOL.English.6.4.a	Identify word origins and derivations.	X
6	SOL.English.6.4.b	Use roots, affixes, synonyms, and antonyms to expand vocabulary.	X
6	SOL.English.6.4.c	Use context and sentence structure to determine meanings and differentiate among multiple meanings of words.	X
6	SOL.English.6.4.d	Identify and analyze the construction and impact of figurative language.	X
6	SOL.English.6.4.e	Use word-reference materials.	X
6	SOL.English.6.4.f	Extend general and cross-curricular vocabulary through speaking, listening, reading, and writing.	X
6	SOL.English.6.5	The student will read and demonstrate comprehension of a variety of fictional texts, literary nonfiction, and poetry.	X
6	SOL.English.6.5.a	Identify the elements of narrative structure, including setting, character, plot, conflict, and theme.	X
6	SOL.English.6.5.b	Describe cause and effect relationships and their impact on plot.	X
6	SOL.English.6.5.c	Explain how an author uses character development to drive conflict and resolution.	X
6	SOL.English.6.5.d	Differentiate between first and third person point-of-view.	X
6	SOL.English.6.5.e	Describe how word choice and imagery contribute to the meaning of a text.	X
6	SOL.English.6.5.f	Draw conclusions and make inferences using the text for support.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
6	SOL.English.6.5.g	Identify the characteristics of a variety of genres.	X
6	SOL.English.6.5.h	Identify and analyze the author’s use of figurative language.	X
6	SOL.English.6.5.i	Compare/contrast details in literary and informational nonfiction texts.	X
6	SOL.English.6.5.j	Identify transitional words and phrases that signal an author’s organizational pattern.	X
6	SOL.English.6.5.k	Use reading strategies to monitor comprehension throughout the reading process.	X
6	SOL.English.6.6	The student will read and demonstrate comprehension of a variety of nonfiction texts.	X
6	SOL.English.6.6.a	Skim materials using text features such as type, headings, and graphics to predict and categorize information.	X
6	SOL.English.6.6.b	Identify main idea.	X
6	SOL.English.6.6.c	Summarize supporting details.	X
6	SOL.English.6.6.d	Create an objective summary including main idea and supporting details.	X
6	SOL.English.6.6.e	Draw conclusions and make inferences based on explicit and implied information.	X
6	SOL.English.6.6.f	Identify the author’s organizational pattern(s).	X
6	SOL.English.6.6.g	Identify transitional words and phrases that signal an author’s organizational pattern.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
6	SOL.English.6.6.h	Differentiate between fact and opinion.	X
6	SOL.English.6.6.i	Identify cause and effect relationships.	X
6	SOL.English.6.6.j	Analyze ideas within and between selections providing textual evidence.	X
6	SOL.English.6.6.k	Use reading strategies to monitor comprehension throughout the reading process.	X
7	SOL.English.7.4	The student will read and determine the meanings of unfamiliar words and phrases within authentic texts.	X
7	SOL.English.7.4.a	Identify word origins and derivations.	X
7	SOL.English.7.4.b	Use roots, affixes, synonyms, and antonyms to expand vocabulary.	X
7	SOL.English.7.4.c	Identify and analyze the construction and impact of figurative language.	X
7	SOL.English.7.4.d	Identify connotations.	X
7	SOL.English.7.4.e	Use context and sentence structure to determine meanings and differentiate among multiple meanings of words.	X
7	SOL.English.7.4.f	Use word-reference materials to determine meanings and etymology.	X
7	SOL.English.7.4.g	Extend general and cross-curricular vocabulary through speaking, listening, reading, and writing.	X
7	SOL.English.7.5	The student will read and demonstrate comprehension of a variety of fictional texts, literary nonfiction, poetry, and drama.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
7	SOL.English.7.5.a	Describe the elements of narrative structure including setting, character development, plot, theme, and conflict and how they influence each other.	×
7	SOL.English.7.5.b	Identify and explain the theme(s).	×
7	SOL.English.7.5.c	Identify cause and effect relationships and their impact on plot.	×
7	SOL.English.7.5.d	Differentiate between first and third person point-of-view.	×
7	SOL.English.7.5.e	Identify elements and characteristics of a variety of genres.	×
7	SOL.English.7.5.f	Compare and contrast various forms and genres of fictional text.	×
7	SOL.English.7.5.g	Describe the impact of word choice, imagery, and literary devices including figurative language in an author's style.	×
7	SOL.English.7.5.h	Compare/contrast details in literary and informational nonfiction texts.	×
7	SOL.English.7.5.i	Make inferences and draw conclusions based on the text.	×
7	SOL.English.7.5.j	Use reading strategies to monitor comprehension throughout the reading process.	×
7	SOL.English.7.6	The student will read and demonstrate comprehension of a variety of nonfiction texts.	×
7	SOL.English.7.6.a	Skim materials using text features including type, headings, and graphics to predict and categorize information.	×
7	SOL.English.7.6.b	Identify an author's organizational pattern using textual clues, such as transitional words and phrases.	×

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
7	SOL.English.7.6.c	Make inferences and draw logical conclusions using explicit and implied textual evidence.	X
7	SOL.English.7.6.d	Differentiate between fact and opinion.	X
7	SOL.English.7.6.e	Identify the source, viewpoint, and purpose of texts.	X
7	SOL.English.7.6.f	Describe how word choice and language structure convey an author's viewpoint.	X
7	SOL.English.7.6.g	Identify the main idea.	X
7	SOL.English.7.6.h	Summarize text identifying supporting details.	X
7	SOL.English.7.6.i	Create an objective summary including main idea and supporting details.	X
7	SOL.English.7.6.j	Identify cause and effect relationships.	X
7	SOL.English.7.6.k	Organize and synthesize information for use in written and other formats.	X
7	SOL.English.7.6.l	Analyze ideas within and between selections providing textual evidence.	X
7	SOL.English.7.6.m	Use reading strategies to monitor comprehension throughout the reading process.	X
8	SOL.English.8.4	The student will apply knowledge of word origins, and figurative language to extend vocabulary development within authentic texts.	X
8	SOL.English.8.4.a	Identify and analyze the construction and impact of an author's use of figurative language.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
8	SOL.English.8.4.b	Use context, structure, and connotations to determine meaning and differentiate among multiple meanings of words and phrases.	×
8	SOL.English.8.4.c	Use roots, affixes, synonyms, and antonyms to determine the meaning(s) of unfamiliar words and technical vocabulary.	×
8	SOL.English.8.4.d	Identify the meaning of common idioms.	×
8	SOL.English.8.4.e	Use word-reference materials to determine meanings and etymology.	×
8	SOL.English.8.4.f	Discriminate between connotative and denotative meanings and interpret the connotation.	×
8	SOL.English.8.4.g	Extend general and cross-curricular vocabulary through speaking, listening, reading, and writing.	×
8	SOL.English.8.5	The student will read and analyze a variety of fictional texts, literary nonfiction, poetry, and drama.	×
8	SOL.English.8.5.a	Analyze how authors' development of characters, conflict, point of view, voice, and tone convey meaning.	×
8	SOL.English.8.5.b	Identify cause and effect relationships and their impact on plot.	×
8	SOL.English.8.5.c	Explain the development of the theme(s).	×
8	SOL.English.8.5.d	Explain the use of symbols and figurative language.	×
8	SOL.English.8.5.e	Make inferences and draw conclusions based on explicit and implied information using references to the text for support.	×
8	SOL.English.8.5.f	Identify and analyze characteristics within a variety of genres.	×

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
8	SOL.English.8.5.g	Compare/contrast details in literary and informational nonfiction texts.	X
8	SOL.English.8.5.h	Compare and contrast the authors' use of word choice, dialogue, form, rhyme, rhythm, and voice in different texts.	X
8	SOL.English.8.5.i	Compare and contrast authors' styles.	X
8	SOL.English.8.5.j	Use reading strategies to monitor comprehension throughout the reading process.	X
8	SOL.English.8.6	The student will read, comprehend, and analyze a variety of nonfiction texts.	X
8	SOL.English.8.6.a	Identify an author's organizational pattern using textual clues, such as transitional words and phrases.	X
8	SOL.English.8.6.b	Apply knowledge of text features and organizational patterns to analyze selections.	X
8	SOL.English.8.6.c	Skim materials to develop an overview or locate information.	X
8	SOL.English.8.6.d	Make inferences and draw conclusions based on explicit and implied information using evidence from text as support.	X
8	SOL.English.8.6.e	Analyze the author's qualifications, viewpoint, word choice, and impact.	X
8	SOL.English.8.6.f	Analyze details for relevance and accuracy.	X
8	SOL.English.8.6.g	Differentiate between fact and opinion.	X
8	SOL.English.8.6.h	Identify the main idea.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
8	SOL.English.8.6.i	Summarize the text identifying supporting details.	X
8	SOL.English.8.6.j	Identify cause and effect relationships.	X
8	SOL.English.8.6.k	Evaluate, organize, and synthesize information for use in written and other formats.	X
8	SOL.English.8.6.l	Analyze ideas within and between selections providing textual evidence.	X
8	SOL.English.8.6.m	Use reading strategies to monitor comprehension throughout the reading process.	X
9	SOL.English.9.3	The student will apply knowledge of word origins, derivations, and figurative language to extend vocabulary development in authentic texts.	X
9	SOL.English.9.3.a	Use structural analysis of roots, affixes, synonyms, and antonyms to understand complex words.	X
9	SOL.English.9.3.b	Use context, structure, and connotations to determine meanings of words and phrases.	X
9	SOL.English.9.3.c	Discriminate between connotative and denotative meanings and interpret the connotation.	X
9	SOL.English.9.3.d	Identify the meaning of common idioms.	X
9	SOL.English.9.3.e	Explain the meaning of literary and classical allusions and figurative language in text.	X
9	SOL.English.9.3.f	Extend general and cross-curricular vocabulary through speaking, listening, reading, and writing.	X
9	SOL.English.9.4	The student will read, comprehend, and analyze a variety of fictional texts including narratives, literary nonfiction, poetry, and drama.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
9	SOL.English.9.4.a	Identify the characteristics that distinguish literary forms.	X
9	SOL.English.9.4.b	Explain the relationships between and among elements of literature: characters, plot, setting, tone, point of view, and theme.	X
9	SOL.English.9.4.c	Interpret how themes are connected across texts.	X
9	SOL.English.9.4.d	Compare and contrast the use of rhyme, rhythm, sound, imagery, and other literary devices to convey a message and elicit the reader's emotion.	X
9	SOL.English.9.4.e	Analyze the cultural or social function of a literary text.	X
9	SOL.English.9.4.f	Explain the relationship between the author's style and literary effect.	X
9	SOL.English.9.4.g	Explain the influence of historical context on the form, style, and point of view of a written work.	X
9	SOL.English.9.4.h	Compare and contrast authors' use of literary elements within a variety of genres.	X
9	SOL.English.9.4.i	Analyze how the author's specific word choices and syntax impact the author's purpose.	X
9	SOL.English.9.4.j	Make inferences and draw conclusions using references from the text(s) for support.	X
9	SOL.English.9.4.k	Compare/contrast details in literary and informational nonfiction texts.	X
9	SOL.English.9.4.l	Use reading strategies to monitor comprehension throughout the reading process.	X
9	SOL.English.9.5	The student will read and analyze a variety of nonfiction texts.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
9	SOL.English.9.5.a	Apply knowledge of text features and organizational patterns to understand, analyze, and gain meaning from texts.	X
9	SOL.English.9.5.b	Make inferences and draw conclusions based on explicit and implied information using evidence from text as support.	X
9	SOL.English.9.5.c	Analyze the author’s qualifications, viewpoint, and impact.	X
9	SOL.English.9.5.d	Recognize an author’s intended purpose for writing and identify the main idea.	X
9	SOL.English.9.5.e	Summarize, paraphrase, and synthesize ideas, while maintaining meaning and a logical sequence of events, within and between texts.	X
9	SOL.English.9.5.f	Identify characteristics of expository, technical, and persuasive texts.	X
9	SOL.English.9.5.g	Identify a position/argument to be confirmed, disproved, or modified.	X
9	SOL.English.9.5.h	Evaluate clarity and accuracy of information.	X
9	SOL.English.9.5.i	Analyze, organize, and synthesize information in order to solve problems, answer questions, complete a task, or create a product.	X
9	SOL.English.9.5.j	Differentiate between fact and opinion and evaluate their impact.	X
9	SOL.English.9.5.k	Analyze ideas within and between selections providing textual evidence.	X
9	SOL.English.9.5.l	Use the reading strategies to monitor comprehension throughout the reading process.	X
10	SOL.English.10.3	The student will apply knowledge of word origins, derivations, and figurative language to extend vocabulary development in authentic texts.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
10	SOL.English.10.3.a	Use structural analysis of roots, affixes, synonyms, and antonyms, to understand complex words.	X
10	SOL.English.10.3.b	Use context, structure, and connotations to determine meanings of words and phrases.	X
10	SOL.English.10.3.c	Discriminate between connotative and denotative meanings and interpret the connotation.	X
10	SOL.English.10.3.d	Explain the meaning of common idioms.	X
10	SOL.English.10.3.e	Explain the meaning of literary and classical allusions and figurative language in text.	X
10	SOL.English.10.3.f	Extend general and cross-curricular vocabulary through speaking, listening, reading, and writing.	X
10	SOL.English.10.4	The student will read, comprehend, and analyze literary texts of different cultures and eras.	X
10	SOL.English.10.4.a	Make inferences and draw conclusions using references from the text(s) for support.	X
10	SOL.English.10.4.b	Analyze the similarities and differences of techniques and literary forms represented in the literature of different cultures and eras.	X
10	SOL.English.10.4.c	Interpret the cultural or social function of world and ethnic literature.	X
10	SOL.English.10.4.d	Analyze universal themes prevalent in the literature of different cultures.	X
10	SOL.English.10.4.e	Examine a literary selection from several critical perspectives.	X
10	SOL.English.10.4.f	Critique how authors use key literary elements to contribute to meaning including, character development, theme, conflict, and archetypes.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
10	SOL.English.10.4.g	Interpret how themes are connected within and across texts.	✕
10	SOL.English.10.4.h	Explain the influence of historical context on the form, style, and point of view of a literary text(s).	✕
10	SOL.English.10.4.i	Evaluate how an author’s specific word choices, syntax, tone, and voice shape the intended meaning of the text.	✕
10	SOL.English.10.4.j	Compare/contrast details in literary and informational nonfiction texts.	✕
10	SOL.English.10.4.k	Compare and contrast how literary devices convey a message and elicit a reader’s emotions.	✕
10	SOL.English.10.4.l	Compare and contrast character development in a play to characterization in other literary forms.	✕
10	SOL.English.10.4.m	Use reading strategies to monitor comprehension throughout the reading process.	✕
10	SOL.English.10.5	The student will read, interpret, analyze, and evaluate nonfiction texts.	✕
10	SOL.English.10.5.a	Analyze text features and organizational patterns to evaluate the meaning of texts.	✕
10	SOL.English.10.5.b	Recognize an author’s intended audience and purpose for writing.	✕
10	SOL.English.10.5.c	Skim materials to develop an overview and locate information.	✕
10	SOL.English.10.5.d	Compare and contrast informational texts for intent and content.	✕
10	SOL.English.10.5.e	Interpret and use data and information in maps, charts, graphs, timelines, tables, and diagrams.	✕

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
10	SOL.English.10.5.f	Draw conclusions and make inferences on explicit and implied information using textual support as evidence.	X
10	SOL.English.10.5.g	Analyze and synthesize information in order to solve problems, answer questions, and generate new knowledge.	X
10	SOL.English.10.5.h	Analyze ideas within and between selections providing textual evidence.	X
10	SOL.English.10.5.i	Summarize, paraphrase, and synthesize ideas, while maintaining meaning and a logical sequence of events, within and between texts.	X
10	SOL.English.10.5.j	Use reading strategies throughout the reading process to monitor comprehension.	X
11	SOL.English.11.3	The student will apply knowledge of word origins, derivations, and figurative language to extend vocabulary development in authentic texts.	X
11	SOL.English.11.3.a	Use structural analysis of roots, affixes, synonyms, and antonyms to understand complex words.	X
11	SOL.English.11.3.b	Use context, structure, and connotations to determine meanings of words and phrases.	X
11	SOL.English.11.3.c	Discriminate between connotative and denotative meanings and interpret the connotation.	X
11	SOL.English.11.3.d	Explain the meaning of common idioms.	X
11	SOL.English.11.3.e	Explain the meaning of literary and classical allusions and figurative language in text.	X
11	SOL.English.11.3.f	Extend general and cross-curricular vocabulary through speaking, listening, reading, and writing.	X
11	SOL.English.11.4	The student will read, comprehend, and analyze relationships among American literature, history, and culture.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
11	SOL.English.11.4.a	Describe contributions of different cultures to the development of American literature.	✕
11	SOL.English.11.4.b	Compare and contrast the development of American literature in its historical context.	✕
11	SOL.English.11.4.c	Analyze American literature, as it reflects traditional and contemporary themes, motifs, universal characters, and genres.	✕
11	SOL.English.11.4.d	Interpret the social or cultural function of American literature.	✕
11	SOL.English.11.4.e	Analyze how context and language structures convey an author's intent and viewpoint.	✕
11	SOL.English.11.4.f	Critique how authors use key literary elements to contribute to meaning including character development, theme, conflict, and archetypes within and across texts.	✕
11	SOL.English.11.4.g	Interpret how the sound and imagery of poetry support the subject, mood, and theme, and appeal to the reader's senses.	✕
11	SOL.English.11.4.h	Evaluate how specific word choices, syntax, tone, and voice support the author's purpose.	✕
11	SOL.English.11.4.i	Analyze the use of dramatic conventions in American literature.	✕
11	SOL.English.11.4.j	Generate and respond logically to literal, inferential, evaluative, synthesizing, and critical thinking questions about the text(s).	✕
11	SOL.English.11.4.k	Compare/contrast literary and informational nonfiction texts.	✕
11	SOL.English.11.5	The student will read, interpret, analyze, and evaluate a variety of nonfiction texts including employment documents and technical writing.	✕
11	SOL.English.11.5.a	Apply information from texts to clarify understanding of concepts.	✕

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
11	SOL.English.11.5.b	Read and correctly interpret an application for employment, workplace documents, or an application for college admission.	X
11	SOL.English.11.5.c	Analyze technical writing for clarity.	X
11	SOL.English.11.5.d	Paraphrase and synthesize ideas within and between texts.	X
11	SOL.English.11.5.e	Draw conclusions and make inferences on explicit and implied information using textual support.	X
11	SOL.English.11.5.f	Analyze multiple texts addressing the same topic to determine how authors reach similar or different conclusions.	X
11	SOL.English.11.5.g	Analyze false premises, claims, counterclaims, and other evidence in persuasive writing.	X
11	SOL.English.11.5.h	Recognize and analyze use of ambiguity, contradiction, paradox, irony, sarcasm, overstatement, and understatement in text.	X
11	SOL.English.11.5.i	Generate and respond logically to literal, inferential, evaluative, synthesizing, and critical thinking questions about the text(s).	X
12	SOL.English.12.3	The student will apply knowledge of word origins, derivations, and figurative language to extend vocabulary development in authentic texts.	X
12	SOL.English.12.3.a	Use structural analysis of roots, affixes, synonyms, and antonyms, to understand complex words.	X
12	SOL.English.12.3.b	Use context, structure, and connotations to determine meanings of words and phrases.	X
12	SOL.English.12.3.c	Discriminate between connotative and denotative meanings and interpret the connotation.	X
12	SOL.English.12.3.d	Explain the meaning of common idioms, and literary and classical allusions in text.	X

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
12	SOL.English.12.3.e	Extend general and cross-curricular vocabulary through speaking, listening, reading, and writing.	×
12	SOL.English.12.4	The student will read, comprehend, and analyze the development of British literature and literature of other cultures.	×
12	SOL.English.12.4.a	Compare and contrast the development of British literature in its historical context.	×
12	SOL.English.12.4.b	Analyze how authors use key literary elements to contribute to meaning and interpret how themes are connected across texts.	×
12	SOL.English.12.4.c	Compare/contrast details in literary and informational nonfiction texts.	×
12	SOL.English.12.4.d	Interpret the social and cultural function of British literature.	×
12	SOL.English.12.4.e	Interpret how the sound and imagery of poetry support the subject, mood, and theme, and appeal to the reader's senses.	×
12	SOL.English.12.4.f	Compare and contrast traditional and contemporary poems from many cultures.	×
12	SOL.English.12.4.g	Evaluate how dramatic conventions contribute to the theme and effect of plays from American, British, and other cultures.	×
12	SOL.English.12.4.h	Use critical thinking to generate and respond logically to literal, inferential, and evaluative questions about the text(s).	×
12	SOL.English.12.5	The student will read, interpret, analyze, and evaluate a variety of nonfiction texts.	×
12	SOL.English.12.5.a	Use critical thinking to generate and respond logically to literal, inferential, and evaluative questions about the text(s).	×
12	SOL.English.12.5.b	Identify and synthesize resources to make decisions, complete tasks, and solve specific problems.	×

Renaissance Star Reading Aligns to the English Standards of Learning for Virginia Public Schools

Standard Grade Level	Standard Code	Standard Description	Star Reading Alignment
12	SOL.English.12.5.c	Analyze multiple texts addressing the same topic to determine how authors reach similar or different conclusions.	×
12	SOL.English.12.5.d	Recognize and analyze use of ambiguity, contradiction, paradox, irony, overstatement, and understatement in text.	×
12	SOL.English.12.5.e	Analyze false premises, claims, counterclaims, and other evidence in persuasive writing.	×

Star Reading alignment

Implementation Guide

Renaissance

See Every Student.

Renaissance Star Implementation Guide

Purposeful. Proven. Powerful. Predictive.

Insight drives the teaching profession. As an educator, you face a steady stream of information, from on-the-fly questioning to formal assessment responses, and you must continually adjust course to help students achieve learning targets. You might consider many factors when deciding how to act on data: curriculum guidelines, pedagogical expertise, district priorities, advice from colleagues, and knowledge of students and school, to name a few. This creates a setting that is ripe for data inquiry—a process that transforms insight into action.

Data inquiry is collaborative, ongoing, and focused on improving instruction and learning. When you engage in data inquiry as part of a “cycle of improvement that involves the regular collection and systematic analysis of evidence,” * you’re poised to routinely examine data, consider what it could mean, and plan for growth. Star Assessments provide data and information that fuel this cycle.

Star Assessments are comprised of Renaissance Star Reading, Renaissance Star Math, Renaissance Star Early Literacy, Renaissance Star Custom, and Renaissance Star CBM. This system of assessments informs data-based decisions and supports instructional frameworks such as Multi-Tier System of Supports (MTSS).

Moreover, Renaissance Star Assessments are accurate, reliable, and valid. They are highly rated for screening by the National Center on Response to Intervention and for progress monitoring by the National Center on Intensive Intervention.

“When data is used as part of an ongoing cycle of improvement... teachers can change their instructional practice to improve student achievement.” *

Let’s get going.

On the following pages, we explain the basics of why and how to administer the Star tests. If you are new to Star Assessments, this information will get you off to a good start. If you have been using Star Assessments for a while, it will help you ensure that you are administering the test with fidelity and explore new ways to enhance your data-inquiry practices.

And remember, we are here to help. Schools that get the most out of Star Assessments take stock of how well they are utilizing Star data and look for ways to improve. We offer a variety of professional learning opportunities to support you in this endeavor. Contact your Renaissance representative or call (800) 338-4204 for information.

Learn more about Star Assessments with Smart Start

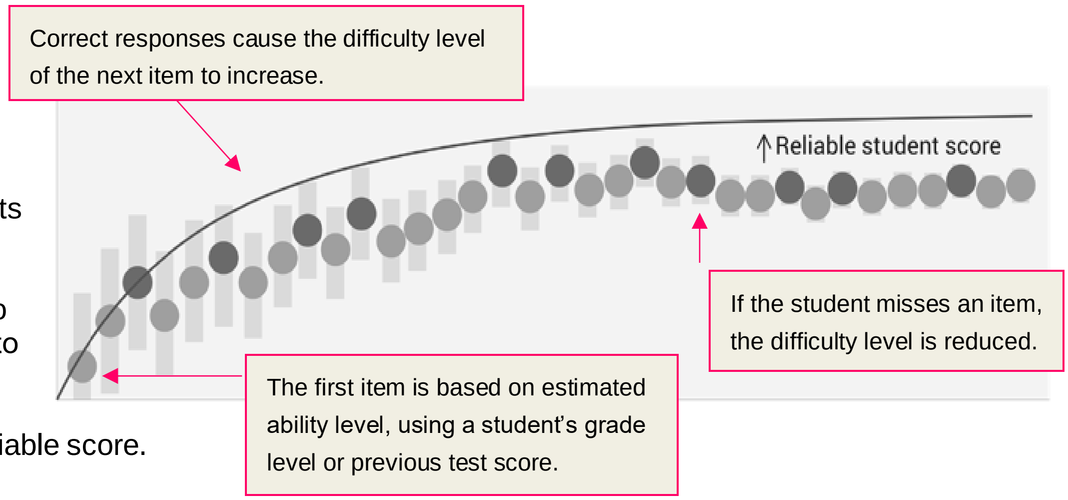
For software instructions, click the question mark in your Star application

*Thessin, Rebecca A. 2015. The need to use evidence in school-based K-12 improvement efforts. Retrieved from http://cep-dc.org /displayDocument.cfm?DocumentID=447.

Getting to know Star Assessments

How the test works.

Star Assessments (Star Reading, Star Math, and Star Early Literacy) are online computer- adaptive tests (CATs). Instead of grade-level test forms, Star tests tailor items to a student's responses to quickly zero in on the student's achievement level and arrive at a reliable score.



Key scores

Scaled score (SS)	Percentile rank (PR)	Student growth percentile (SGP)
Is based on the difficulty of items and the number of correct responses. It is useful for comparing performance across grades. All norm-referenced scores are derived from the scaled score. Scaled scores on the Unified Scale range from 0–1400 in Star Reading, Star Math, and Star Early Literacy.	Indicates the percentage of students nationally who obtained a scaled score equal to or lower than the score of a student. PRs are norm-referenced scores and range from 1–99. A student with a PR of 75 performed as well as or better than 75% of same-grade students nationwide.	Compares a student's growth from one period to the next with that of his or her academic peers nationwide—same-grade students with a similar scaled score history. SGPs range from 1–99: lower numbers show lower relative growth; higher numbers indicate higher relative growth. A student with an SGP of 35 grew more quickly than 35% of academic peers.

Test design

While each Star test is individualized and unique, blueprints ensure that a certain number of items from the domains and skill sets are presented to each student.

	Renaissance Star Reading	Renaissance Star Math	Renaissance Star Early Literacy
For whom is the test designed?	Grades K-12 For students who have basic reading skills.	Grades K-12 For students who have basic reading and math skills.	Pre-K-grade 3 For beginning readers who do not yet read independently or need early literacy skills assessed.
What content is on the test?	Word Knowledge and Skills Comprehension Strategies and Consructed Meaning Analyzing Literary Text Understanding Author's Craft Analyzing Argument and Evaluating Text	Numbers Operations Algebra Geometry Measurement Data Analysis, Statistics, and Probability	Alphabetic Principle Concept of Word Visual Discrimination Phonemic Awareness Phonics Structural Analysis Vocabulary Sentence-Level Comprehension Paragraph-Level Comprehension Early Numeracy

The testing experience

1. Students log in with a user name and password. They test on desktops, laptops, or tablets seven inches or greater.

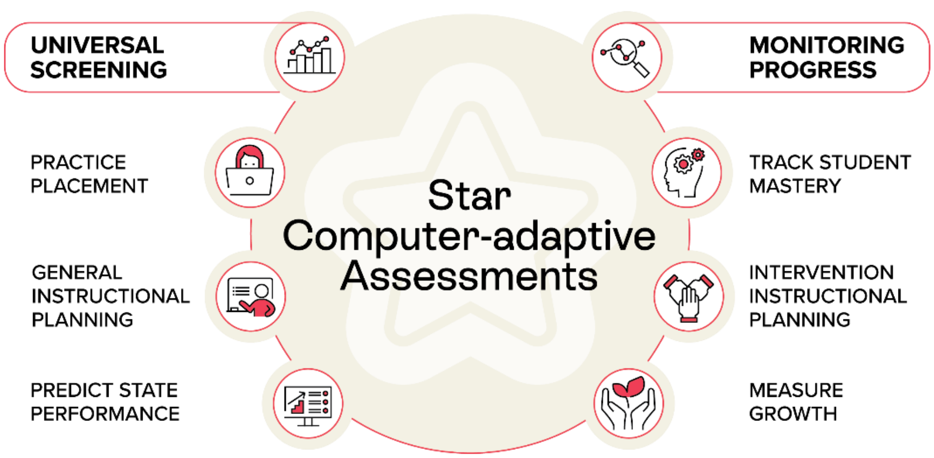
2. As students test, the software adjusts the difficulty of each item. Students answer 34 items for Star Reading and Star Math (or 27 items for Star Early Literacy).

3. After students test, you have access to the results. Star tests take about 15-20 minutes on average. You can view data through a variety of dashboards and reports.

How the test supports data inquiry.
Star Assessments provide a wealth of actionable data. Information from Star Assessments helps you see which students are gaining ground or falling behind, where to focus instruction, who may require intervention, and whether your curriculum and interventions are making a difference. When deciding how Star data will help you answer questions about student performance, keep in mind the following.

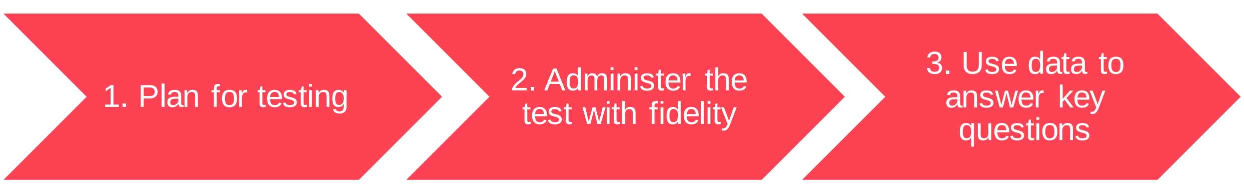
Data from one testing event can be used in multiple ways**.** Suppose you screen all students in the fall, winter, and spring to get a baseline and see how students' progress. The same data can be used to help you answer additional questions: Where should you focus instruction for a student, group, or class? How are students performing in relation to standards? Are students on track to reach proficiency on the state test? You may even use screening data as part of progress-monitoring for some students.

Identify expectations to gauge whether students are meeting them. Setting expectations for student growth, achievement, and standards mastery enables you to view student performance in relation to them. How quickly should students grow? How much do you expect them to achieve by the end of the school year? Your school, district, or state, as well as individual student needs, may also influence expectations.



Three Steps for Acting on Data

The three steps described in this section will help you get started with your Star implementation. The process begins with planning, including scheduling testing (e.g., universal screenings) and preparing to interpret and act on student data. Next, you will be ready to gather data; we explain the critical step of how to administer the test with fidelity. Finally, you will use the data to help answer key questions about student performance.

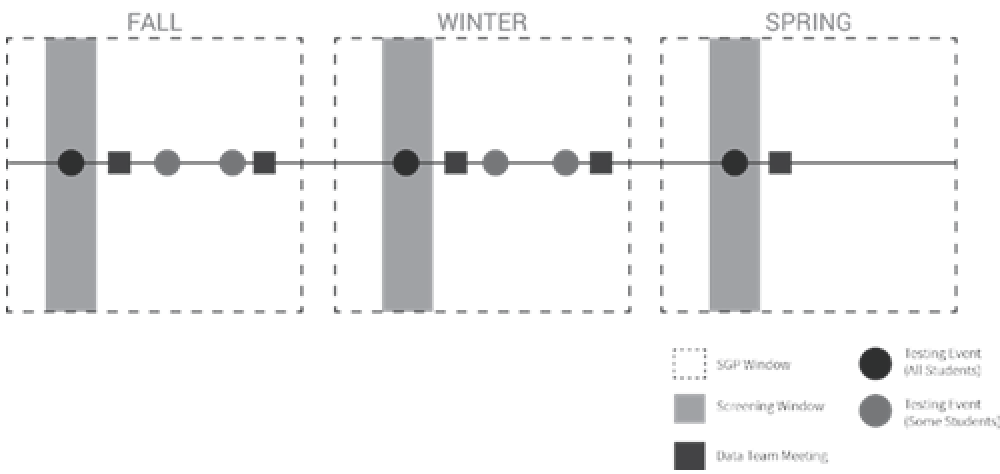


Step 1: Plan for testing.

Planning helps you get the most out of Star Assessments. Consider when and how often students will test, schedule opportunities for reviewing and discussing data, and determine how benchmarks will help you interpret student performance.

Consider frequency and timing

In the example testing schedule below, all students test early in the fall, winter, and spring within the screening windows. (Star software has default screening windows in September, January, and May. The dates can be changed, and more screening dates can be added, up to a maximum of 10 for the school year.) Each screening event is followed by a data team meeting, helping teachers and administrators act on the data quickly. Some students are identified for progress monitoring and test between screenings. Data teams meet periodically during the school year to monitor growth and adjust instructional plans.



Establish data teams.

Data teams take many forms, but they are often comprised of teachers and administrators. Data teams meet to review data from multiple sources and discuss how it can be used to improve instruction and student outcomes. Data team conversations lead to group ownership and responsibility, set the stage for improved data literacy, and promote the emergence of a data culture. Data team meetings vary in frequency; they typically occur after major testing events, such as screenings, and as needed to support instructional decisions.

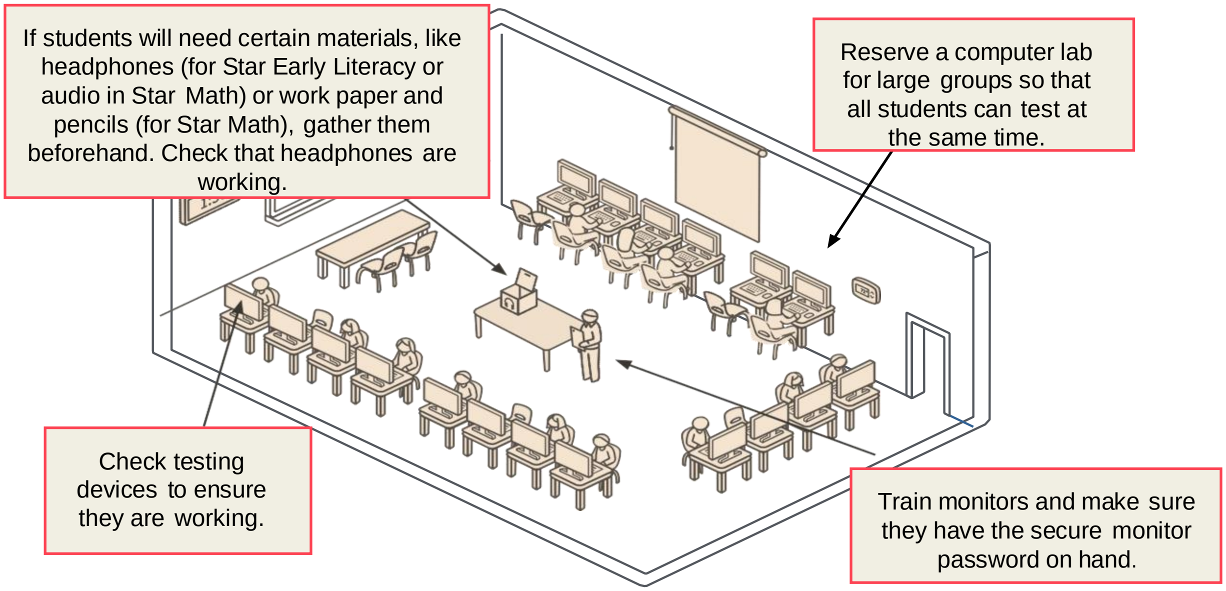
Define benchmarks.

When measuring student performance, it helps to have a benchmark in mind—the lowest level of achievement considered acceptable. A benchmark provides context to help you interpret student performance. In Star software, you can select a school, district, or state benchmark. Choose the one that will best inform your decision-making process. Toggle between benchmarks for alternate views.

Step 2: Administer the test with fidelity.

Because you will make important instructional decisions based, in part, on Star data, it is critical that you maintain the integrity of that data. You can do this by replicating the norming conditions as closely as possible and by following test protocol to give all students the same chance to do their best. Administering the test with fidelity helps ensure that scores reflect students' level of achievement.

Create a good testing environment



Administer the test

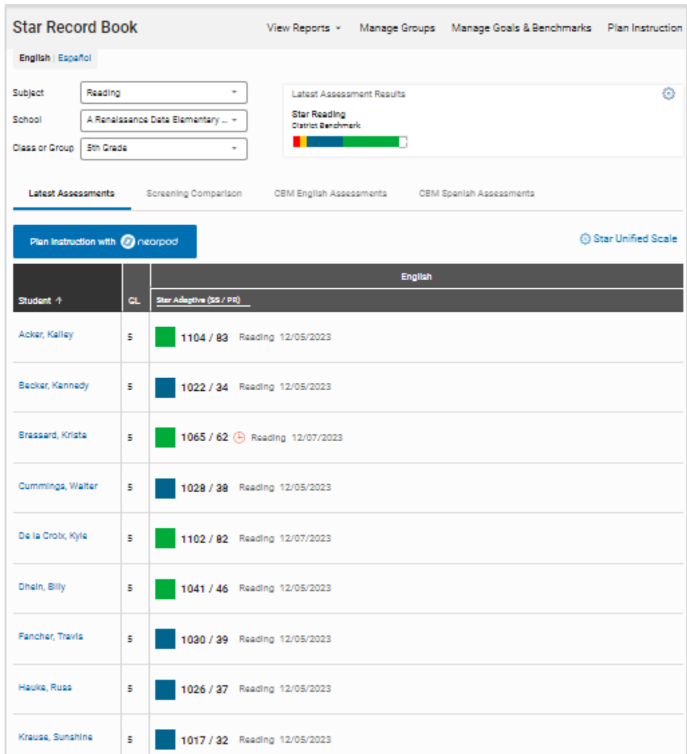
The high-level checklist below lists some main tasks required for administering the Star tests. Mark the steps as you complete them; remaining tasks may help you strengthen your implementation. Refer to the checklist throughout the school year to guide discussions with your colleagues.

Get your schedule ready	Get your environment ready	Get your students ready
☐ Plan universal screenings (e.g., fall, winter, and spring). ☐ Schedule an appropriate time and place to accommodate all students' testing needs. ☐ Train and schedule monitors for the event. ☐ Identify students who require adaptations to the testing experience.	☐ Check testing devices to ensure they are working. ☐ Gather necessary materials and make them available to students. ☐ Make sure test monitors have the monitor password (if set in the software).	☐ Go through the pretest instructions. ☐ Distribute usernames and passwords to students.

Step 2: Use data to answer key questions.

You have assessed students, defined expectations for achievement and growth, and assembled a team for data inquiry. What are your key questions about student performance? How can data from Star Assessments help you answer them? We explore a few possibilities in this section.

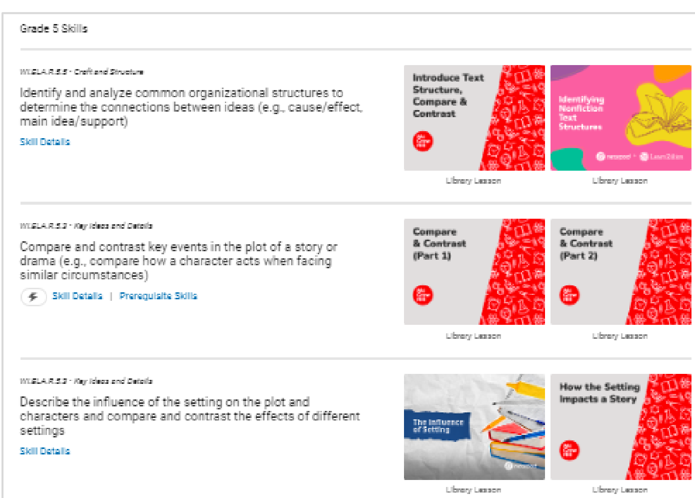
How are students starting out the school year?



ar **Record Book** to who tested and when.

t's Scaled Score, anks and how they relation to grade-level t benchmarks.

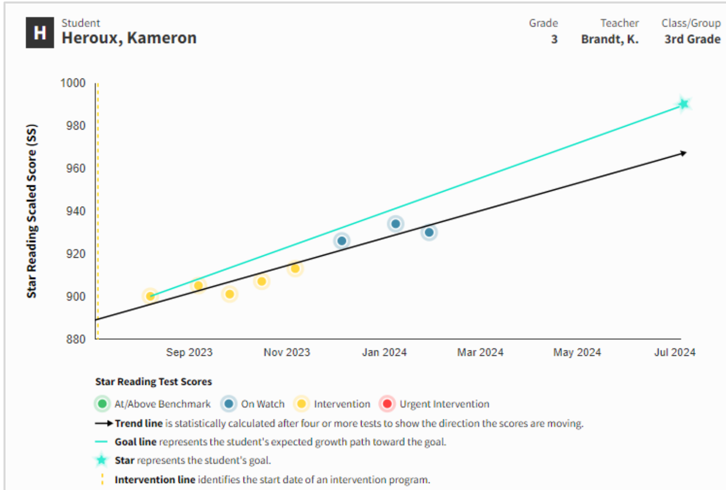
How can I help students grow?



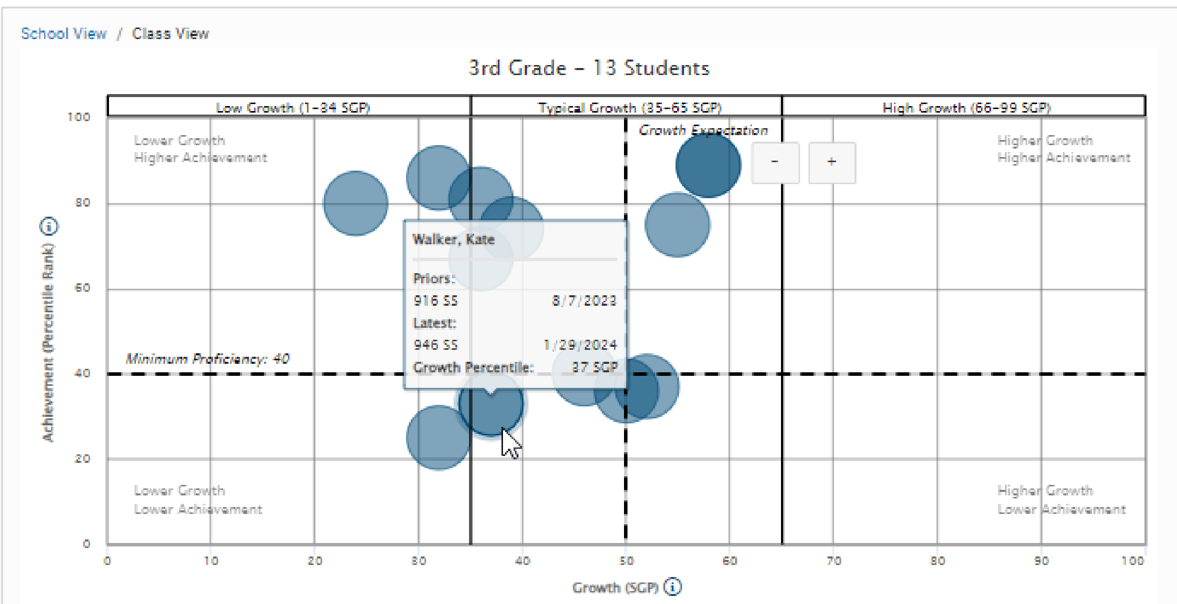
Select **Plan Instruction with Nearpod** in the Record Book to view suggested skills for students or groups of students and access Nearpod lessons. **Skill details** gives you more information on the suggested skill like related standards and prerequisite skills.

Are students on track to meet growth and achievement expectations?

After setting a goal for a student in intervention, use the Student Progress Monitoring Report to track progress toward the goal. Determine whether the student is responding to the intervention.



The **Star Growth Proficiency Chart** displays each student's growth using (SGP) and achievement in one chart.



If you ask...	Check the...	So you can...
How did students do on the test?	Record Book	View the most recent scaled score for each student, the corresponding benchmark category, and the test date.
Where can I find more detail about a student's testing experience?	Diagnostic Report	Access key scores, along with definitions, for a single test event. Also check how much time a student took to complete the test.
How well are students mastering standards and skills?	State Standards Mastery Report	See the levels of mastery that students have achieved in relation to your state standards, as well as a prediction of what students' mastery levels will be at the end of the school year.
How quickly are students growing?	Growth Report	View student SGP's to see precisely how much students are growing, showcasing changes in scores between testing events.

Do you have more questions about student growth and achievement? Star Assessments' array of data displays can help you gain insight as you go about answering them, perhaps with the help of a data team. The reports and dashboards in this section are a starting point; be sure to explore the software to become familiar with all your options. Also check in with colleagues to share your data-inquiry goals and gather input for how Star data can help you achieve them.

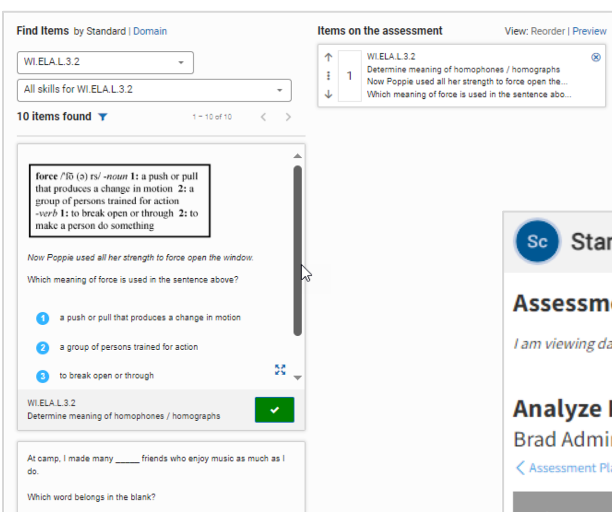
Customize assessments for focused instruction with Star Custom

Star Custom is a formative assessment tool that allows you to administer fixed-form assessments to target specific domains, standards, and skills. Choose skills for assessment and instruction using our data-based learning progressions. Star Custom assessments can be created, scheduled, and sent directly to students. The feedback you receive on Star Custom assessments will help provide unique insight into how students are performing related to targeted skills, grade-level standards, or district expectations.

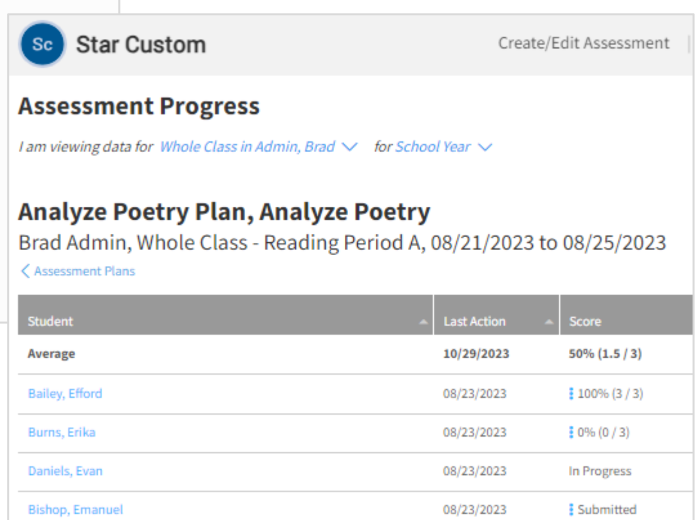
Gauge lesson effectiveness. Assign a pretest to check which upcoming skills students already know. Schedule a posttest to gauge whether students mastered the material. Identify strengths and weaknesses of your lessons from the results.

Eliminate barriers to learning. View item response reports to see common misunderstandings students have about learning material. Explicitly teach with strategies that guide students away from these barriers to learning.

Probe mastery of specific skills. Target a subset of grade-level skills to check the degree of students' mastery. Determine where learning gaps exist, which students require more instruction, and which students are ready to move on.



Discover premade items by filtering by standards and skills to build your own assessments.



The **Assessment Progress** page shows student results in multiple views.

Enhance your view of student development with Star CBM

Star CBM (Curriculum Based Measures) is a Star assessment for students in Kindergarten through 6th grade in Reading, and Kindergarten through 3rd grade in Math. Whereas Star computer-adaptive tests (CATs) presents easier or more difficult items based on student performance during the test, Star CBM uses fixed forms to measure and monitor student performance, via a 1:1 administration by the teacher.

Like Star CATs, Star CBM is a General Outcome Measure (GOM), showing student performance as it grows over time to reach important academic outcomes. With up to 20 forms for each measure of Star CBM, teachers can personally monitor students throughout the year as they grow and progress towards their long-term academic goals. Star CBM allows you to directly assess that growth so you can better tailor instruction and intervention to each learner's specific needs.

Gain insights into your students' academic growth.

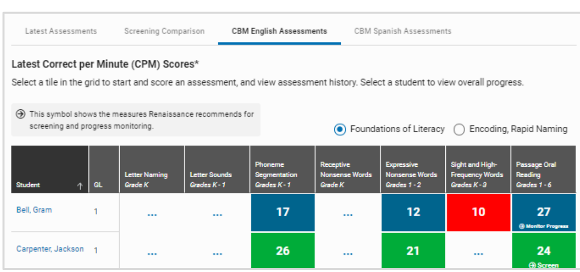
Efficient, 1-to-1 Assessment. Observe your students' abilities in real time as you test students on paper, online (including remotely), or in a mixed format. As you note each child's errors, you can gain immediate insight into areas where students need further support. Each measure takes just one minute to administer.

Screen and monitor progress. A single screening measure is recommended per grade, and the individual student results will prompt the teacher to consider setting a goal and progress monitoring.

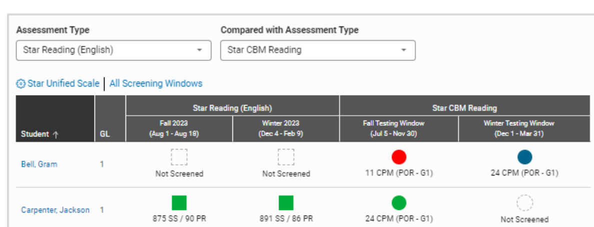
Determine next steps in instructional planning. Evaluate MTSS intervention effectiveness through data. Make decisions on whether instructional adjustments are needed, or further measures should be tested.

Analyze CAT and CBM data side by side. Review trend data from student CBM responses to pinpoint specific challenges for individual students. Analyze Star CBM and Star CAT data side by side to quickly see how students are progressing or note inconsistencies.

In the CBM's Star Record Book, see how students are progressing through CPM (Correct Per Minute) measure scores.



View CBM and CAT tests side by side for easy comparison of results. Look at benchmarks as well as risk categories for each student.



Star Math and Star Reading Technical Manuals

Renaissance

Star Assessments™ for Math Technical Manual

Renaissance
Star Math

Renaissance Learning
PO Box 8036
Wisconsin Rapids, WI 54495-8036
Telephone: (800) 338-4204
(715) 424-3636
Outside the US: 1.715.424.3636
Fax: (715) 424-4242
Email (general/technical questions): support@renaissance.com
Email (international support): worldsupport@renaissance.com
Website: www.renaissance.com

Copyright Notice

8/2023 SM

Contents

Introduction

Star Math: Screening and Progress-Monitoring Assessment

Since the 2011–2012 school year, two different versions of Star Math have been available for use in assessing the mathematical abilities of students in grades K–12. The comprehensive version is a 34-item standards-based adaptive assessment, aligned to state and national curriculum standards, that takes an average of less than 25 minutes. A shorter, 24-item version takes an average of less than 14 minutes, making it a popular choice for progress monitoring in programs such as Response to Intervention. Both versions provide immediate feedback to teachers and administrators on each student's mathematical ability.

Star Math Purpose

As a periodic progress-monitoring assessment, Star Math progress monitoring serves three purposes. First, it provides educators with quick and accurate estimates of students' instructional math levels. Second, it assesses math levels relative to national norms. Third, it provides the means for tracking growth in a consistent manner longitudinally for all students. This is especially helpful to school- and district-level administrators.

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

The Star Math test's repeatability and flexible administration provide specific advantages for everyone responsible for the education process:

- For students, Star Math software provides a challenging, interactive, and brief test that builds confidence in their math ability.
- For teachers, the Star Math test facilitates individualized instruction by identifying children who need remediation or enrichment most.
- For principals, the Star Math software provides regular, accurate reports on performance at the class, grade, and building level.
- For district administrators and assessment specialists, it provides a wealth of reliable and timely data on math growth at each school and districtwide. It also provides a valid basis for comparing data across schools, grades, and special student populations.

This manual documents the suitability of Star Math computer-adaptive testing for these purposes and demonstrates quantitatively how well this innovative instrument in math assessment performs.

Star Math is similar in many ways to the Star Math progress monitoring version, but with some enhanced features, including additional reports and expanded benchmark management.

Design of Star Math

Two Generations of Star Math Assessments

The introduction of the current version of Star Math in 2011 marked the second generation of Star Math assessments. The first generation consisted of the Star Math Progress Monitoring version, which is a fixed-length 24-item adaptive assessment of math levels. This original version of Star Math was published in 1998 and used Item Response Theory (IRT) as the psychometric foundation for adaptive item selection and scoring. Star Math's original item bank contained 2,000+ items spanning more than 200 objectives.

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

The Star Math Progress Monitoring test is further divided into two parts. The first part of the test, the first sixteen items, includes items only from the Numeration

Concepts and the Computation Processes strands. The first eight test items (items 1–8) are from the Numeration Concepts strand, and the following eight test items (items 9–16) are from the Computation Processes strand.

The second part of the test, or the final eight items, includes items from all of the remaining strands. Hence, items 17–24 are drawn from the following six strands: Word Problems; Estimation; Data Analysis and Statistics; Geometry; Measurement; and Algebra. The specific makeup of the strands used in the final eight items depends on the student's grade level. For example, a student in grade 1 will not receive items from the Estimation strand, but items from this strand could be administered to a student in grade 12.

The decision to weight the test heavily toward Numeration Concepts and Computation Processes resulted from the fact that these strands are fundamental to all others, and they include the content about which teachers desire the most information. Although this approach emphasizes the two strands in the first part of the test, it provides adequate content balance to assure valid assessment. Additionally, factor analysis of the various content strands supports the fundamental unidimensionality of the construct being measured in the Star Math Progress Monitoring test.

The second generation is the current version of Star Math published in 2011. This is the first version of Star Math to be designed as a standards-based test. The organization of the content in Star Math differs from that of the original Star Math test—the Star Math Progress Monitoring test. Star Math's content organization reflects current thinking, as embodied in many different sets of national and local curriculum standards. The following four domains were identified and included in Star Math: Numbers and Operations; Algebra; Geometry & Measurement; and Data Analysis, Statistics & Probability. Within each of these domains, skills are organized into skill sets; there are 54 skill sets in all, comprising a total of over 790 core skills.

The Star Math test is a 34-item standards-based version, administered as 6 blocks of items in a single section. Each block of items contains a blend of items from the 4 domains. The number of items administered in a block varies by grade band. The item sequencing calls for more content balance at the beginning, middle, and end of the test by "spiraling" the content throughout the test, thus ensuring that the math ability estimate at any point during a test is based on a broad range of content, rather than on a limited sample of skills.

Thus, this second generation differed from the first in three major respects: It organized the content differently, its test length increased to 34 items, and the size of the item banks grew to over 6,000 items. Like the first generation of Star Math tests, the second generation continues to measure a single construct: mathematical achievement.

Overarching Design Considerations

One of the fundamental Star Math design decisions involved the choice of how to administer the test. The primary advantage of using computer software to administer Star Math tests is the ability to tailor each student's test based on his or her responses to previous items. Conventional assessments, including paper-and-pencil tests, typically entail fixed test forms: every student must respond to the same items in the same sequence. Using computer-adaptive procedures, it is possible for students to test on items that appropriately match their current level of proficiency. The item selection procedures, termed Adaptive Branching, effectively customize the test for each student's achievement level.

Adaptive Branching offers significant advantages in terms of test reliability, testing time, and student motivation. Reliability improves over fixed-form tests because the test difficulty is adjusted to each individual's performance level; students do not have to fit a "one test fits all" model. Most of the test items that students respond to are at levels of difficulty that closely match their achievement level. Testing time decreases because, unlike in paper-and-pencil tests, there is no need to expose every student to a broad range of material, portions of which are inappropriate because they are either too easy for high achievers or too difficult for those with low current levels of performance. Finally, student motivation improves simply because of these issues—test time is minimized and test content is neither too difficult nor too easy.

Another fundamental Star Math design decision involved the choice of the content and format of items for the test. Many types of stimulus and response procedures were explored, researched, discussed, and prototyped. The traditional multiple-choice format was chosen. This decision was made for interrelated reasons of efficiency, breadth of construct coverage, and objectivity and simplicity of scoring.

In both Star Math Progress Monitoring and Star Math, all management and test administration functions are controlled using a management system which is accessed by means of a computer with web access. This makes a number of features possible:

- It makes it possible for multiple schools to share a central database, such as a district-level database. Records of students transferring between schools within the district will be maintained in the database; the only information that needs revision following a transfer is the student's updated school and class assignments.
- The same database that contains Star Math data can contain data on other Star tests, including Star Early Literacy and Star Reading. The Renaissance program is a powerful information management program that allows you to manage all your district, school, personnel, and student data in one place. Changes made

to district, school, teacher, and student data for any of these products, as well as other Renaissance software, are reflected in every other Renaissance program sharing the central database.

- Multiple levels of access are available, from the test administrator within a school or classroom to teachers, principals, and district administrators.
- Renaissance takes reporting to a new level. Not only can you generate reports from the student level all the way up to the school level, but you can also limit reports to specific groups, subgroups, and combinations of subgroups. This supports "disaggregated" reporting; for example, a report might be specific to students eligible for free or reduced lunch, to English language learners, or to students who fit both categories. It also supports compiling reports by teacher, class, school, grade within a school, and many other criteria such as a specific date range. In addition, the Renaissance consolidated reports allow you to gather data from more than one program (such as Star Math and Accelerated Math) at the teacher, class, school, and district level and display the information in one report.
- Since the Renaissance software is accessed through a web browser, teachers (and administrators) will be able to access the program from home.
- For both versions of Star Math, all shortcuts to the student program will automatically redirect to the browser-based program (the Renaissance Welcome page) each time they are used.

Test Interface

The Star Math test interface was designed to be both simple and effective. Students can use either the mouse or the keyboard to answer questions.

- If using the keyboard, students press one of the four letter keys (**A, B, C,** and **D**) and then press the **Enter** key (or the **return** key on Macintosh computers).
- If using the mouse, students click the answer of choice and then click **Next** to enter the answer.
- On a tablet, students tap their answer choice; then, they tap **Next**.

Practice Session

Star Math software includes a provision for a brief practice test preceding the test itself. The practice session allows students to get comfortable with the test interface and to make sure that they know how to operate it properly. As soon as a student has answered two out of three practice questions correctly, the program takes the student into the actual test. If the student has not successfully answered two of the three items by the end of the practice session, Star Math will present three more questions, and the student can pass the practice session by answering two of those questions correctly. If the student does not pass after the second attempt, the student will not proceed to the actual Star Math test. Even students with low math and reading skills should be able to answer the practice questions correctly. However, Star Math will halt the testing session and tell the student to ask the teacher for help if the student does not pass the practice session after the second attempt.

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

Once a student has successfully passed a practice session, the student will not be presented with practice items again on a test of the same type taken within the next 180 days.

Adaptive Branching/Test Length

Star Math’s branching control uses a proprietary approach somewhat more complex than the maximum information criterion based on the Rasch model. The Star Math approach was designed to yield reliable test results for both the criterion-referenced and norm-referenced scores by adjusting item difficulty to the responses of the individual being tested while striving to minimize test length and student frustration.

In order to minimize student frustration, the first administration of the Star Math test begins with items that have a difficulty level that is below what a typical student at a given grade can handle—usually one or two grades below grade placement. On the average, about 85 percent of students will be able to answer the first item correctly. Teachers can override the use of grade placement for determining starting difficulty by entering the current level of mathematics instruction for the student using the MIL (Math Instructional Level). When an MIL is provided, the program uses that value to raise or lower the starting difficulty of the first test. On the second and subsequent administrations, the test begins about one grade lower than the ability last demonstrated within 180 days. Students generally have an 85 percent chance of answering the first item correctly on second and subsequent tests.

Test Length

Once the testing session is underway, the Star Math test administers 34 items (or 24 items for the Star Math Progress Monitoring test) of varying difficulty based on the student's responses; this is sufficient information to obtain a reliable Scaled Score and to determine the student's math Level.

The length of time needed to complete a Star Math test varies across students.

Table 1 provides an overview of the testing time by grade for the students who took the full-length 34-item version of Star Math during the 2018–2019 school year. The results of the analysis of test completion time indicate that half or more of students completed the test in less than 25 minutes, depending on grade, and even in the slowest grade (grade 6) 95% of students finished their Star Math test in less than 42 minutes.

Table 1: Average and Percentiles of Total Time to Complete the 34-item Star Math Assessment During the 2018–2019 School Year

Grade	Sample Size	Time to Complete Test (in Minutes)					
		Mean	Standard Deviation	5th Percentile	50th Percentile	95th Percentile	99th Percentile
K	88,566	13.24	5.67	8.28	11.62	23.45	34.00
1	1,381,713	15.16	6.21	8.58	13.62	26.85	35.85
2	2,010,310	17.90	7.12	9.27	16.52	31.13	40.22
3	2,110,198	22.32	8.58	10.80	21.07	37.95	47.73
4	2,093,035	23.85	8.69	11.85	22.68	39.57	49.30
5	2,061,995	24.59	8.58	12.50	23.53	40.10	49.23
6	1,685,463	25.85	8.69	13.12	25.02	41.27	49.92
7	1,413,980	25.53	8.51	12.82	24.82	40.40	48.53
8	1,339,122	24.82	8.27	12.50	24.10	39.32	47.18
9	601,062	22.97	8.51	10.85	21.98	38.25	46.57
10	443,006	22.32	8.44	10.43	21.32	37.48	45.77
11	278,229	21.74	8.53	10.03	20.60	37.23	45.77
12	154,255	21.01	8.68	9.57	19.63	36.97	46.10

Table 2 provides an overview of the Star Math Progress Monitoring testing time by grade for the students using data from the 2017–2018 and 2018–2019 school years. For that version of the test, about half of the students at every grade completed the Star Math Progress Monitoring test in less than 13 minutes, and even in the slowest grade (grade 4) 95% of students finished in less than 23 minutes.

Table 2: Average and Percentiles of Total Time to Complete the 24-item Star Math Progress Monitoring Assessment During the 2017–2018 and 2018–2019 School Years

Grade	Sample Size	Time to Complete Test (in Minutes)					
		Mean	Standard Deviation	5th Percentile	50th Percentile	95th Percentile	99th Percentile
1	9,239	10.14	4.25	5.45	9.05	18.50	23.90
2	17,873	11.19	4.60	5.67	10.22	20.05	25.53
3	16,445	12.86	5.20	6.15	11.95	22.53	28.28
4	18,372	13.40	5.08	6.60	12.65	22.92	28.48
5	15,733	13.93	5.05	6.93	13.22	23.18	28.68
6	10,390	13.86	5.04	6.78	13.22	23.13	28.55
7	8,741	13.57	4.83	6.67	13.05	22.32	26.85
8	7,699	12.82	4.75	6.32	12.13	21.73	26.37
9	875	10.95	4.89	5.68	9.85	19.70	26.28
10	731	10.70	4.46	5.55	9.63	18.85	25.78
11	518	10.73	4.10	5.83	9.89	18.63	23.30
12	362	10.62	4.38	5.68	9.49	19.78	26.98

Test Repetition

Star Math score data can be used for multiple purposes such as screening, placement, planning instruction, benchmarking, and outcomes measurement. The frequency with which the assessment is administered depends on the purpose for assessment and how the data will be used. Renaissance Learning recommends assessing students only as frequently as necessary to get the data needed. Schools that use Star for screening purposes typically administer it two to five times per year. Teachers who want to monitor student progress more closely or use the data for instructional planning may use it more frequently. Star Math may be administered monthly for progress monitoring purposes, and as often as weekly when needed.

Star Math keeps track of the questions presented to each student from test session to test session and will not ask the same question more than once in any 120-day period.

Item Time Limits

The Star Math tests place no limits on total testing time. However, there are time limits for each test item. The per-item time limits are generous and ensure that more than 90% of students can complete each item within the normal time limits. Each practice question has a 90-second time limit and each test question has a 4-minute time limit.

Standard Time Limits:

- Practice questions: 90 seconds (1.5 minutes) for each question
- Test questions 240 seconds (4 minutes) for each question

Star Math also provides the option of extended time limits for selected students who, in the judgment of the test administrator, require more than the standard amount of time to read and answer the test questions. Extended time limits are twice as long as standard time limits.

Extended Time Limits:

- Practice questions: 180 seconds (3 minutes) for each question
- Test questions: 480 seconds (8 minutes) for each question

Extended time may be a valuable accommodation for English language learners as well as for some students with disabilities. Test users who elect the extended time limit for their students should be aware that Star Math norms, as well as other technical data such as reliability and validity, are based on test administration using the standard time limits. When the extended time limit accommodation is elected, students have two times longer than the standard time limits to answer each question.

At all grades, regardless of the extended time limit setting, when a student has only 15 seconds remaining for a given item, a time-out warning appears, indicating that he or she should make a final selection and move on. Items that time out are counted as incorrect responses unless the student has the correct answer selected when the item times out. If the correct answer is selected at that time, the item will be counted as a correct response.

If a student doesn't respond to an item, the item times out and briefly gives the student a message describing what has happened. Then the next item is presented. The student does not have an opportunity to take the item again. If a student doesn't respond to any item, all items are scored as incorrect.

Accessibility and Test Accommodations

The Star Math test can be accessed in an accessible format that is in compliance with WCAG 2.1 AA. This format allows for users with different ability levels to access the test utilizing different modalities, including assistive technology such as the JAWS screen reader. The content of the item bank is nearly identical to the traditional item delivery format, with the user interface modified slightly and a small number of visually biased items removed from the item bank. A student will be presented with the WCAG 2.0 AA version of the test after educators select one of the relevant test accommodations available in that student's Personal Needs

Profile. Some of the available accommodations are the ability to change the size of the text or the color contrast, a highlighter, a line reader, an answer choice eliminator, a calculator or unlimited time to answer questions. In order to provide the best experience for students and teachers, the available accommodations could be modified during the school year.

Unlimited Time

Beginning with the 2022–23 school year, a new preference has been added: the Accommodations Preference. Among other things, this preference allows teachers to give students virtually unlimited time to answer questions: 15 minutes for both practice questions and test questions. When this preference is set, the student will not see a time-out warning when there are 15 seconds left; however, if there is no activity at all from the student within 15 minutes of a question first being presented, the student will be shown a dialog box. The student will have 60 seconds to close the dialog box and return to the test. If the student does not close the dialog box within 60 seconds, the student's current progress on the test will be saved and the test will be ended (and can be resumed the same way as a paused test).

Test Security

Star Math software includes a number of security features to protect the content of the test and to maintain the confidentiality of the test results.

Split Application Model

When students log into Star Math, they do not have access to the same functions that teachers, administrators, and other personnel can access. Students are allowed to take the test, but no other features available in Star Math are available to them; therefore, they have no access to confidential information. When teachers and administrators log in, they can manage student and class information, set preferences, and create informative reports about student test performance.

Individualized Tests

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

Data Encryption

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

Access Levels and Capabilities

Each user's level of access to a Renaissance program depends on the primary position assigned to that user. Each primary position is part of a user permission group. There are six of these groups: district level administrator, district dashboard owner, district staff, school level administrator, school staff, and teacher. By default, each user permission group is granted a specific set of user permissions; each user permission corresponds to one or more tasks that can be performed in the program. The user permissions for these groups can be changed, and user permissions can be granted or removed on an individual level.

Renaissance also allows you to restrict students' access to certain computers. This prevents students from taking Star Math tests from unauthorized computers (such as home computers). For more information on student access security, see https://star-help.renaissance.com/hc/en-us/articles/11945235255195.

The security of the Star Math data is also protected by each person's username (which must be unique) and password. Usernames and passwords identify users, and the program only allows them access to the data and features that they are allowed based on their position and the user permissions that they have been granted. Personnel who log in to Renaissance (teachers, administrators, or staff) must enter a username and password before they can access the data and create reports. Without an appropriate username and password, personnel cannot use the Star Math software.

Test Monitoring/Password Entry

Test monitoring is another useful Star Math security feature. Test monitoring is implemented using the Password Requirement preference, which specifies whether monitors must enter their passwords at the start of a test. Students are required to enter a username and password to log in before taking a test. This ensures that students cannot take tests using other students' names.

Final Caveat

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

Test Administration Procedures

In order to ensure consistency and comparability of results to the Star Math norms, students taking Star Math tests should follow standard administration procedures. The testing environment should be as free from distractions for the student as possible.

The Test Administration Manual included with the Star Math product (https://docs.renaissance.com/R63824) describes the standard test orientation procedures that teachers should follow to prepare their students for the Star Math test. These instructions are intended for use with students of all ages and were successfully field-tested with students ranging from grades 1–12. It is important to use these same instructions with all students before they take the Star Math test.

Content and Item Development

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

Content Specification: Star Math

Since the introduction of the initial version of the Star Math test in 1998, it has undergone a process of continuous research and improvement, and has evolved into the two distinct versions now in use. The Star Math Progress Monitoring version is the direct descendant of Star Math version 1: a 24-item test of general math achievement based on content that is heavily weighted towards numeration concepts and operations. Star Math itself is now a 34-item standards-based assessment, with a content distribution that changes as grade levels increase between the primary and high school grades.

Relative to Star Math Progress Monitoring, Star Math is an expanded test with new content and several technical innovations. The Star Math item bank has expanded from the original bank of 1,900 test items to more than 6,200 test items and will continue to grow as standards and curriculums evolve. The Star Math test content began with 210 skills and has expanded to include 790 skills that significantly enhance the test's ability to measure math skills in various state learning progressions.

For information regarding the development of Star Math items, see "Item Development Guidelines: Star Math" on page 5. Before inclusion in the Star Math item bank, all Star Math items are reviewed to ensure they meet the content specifications for Star Math item development. Items that do not meet the specifications are revised and recalibrated or discarded. All new item development adheres to the content specifications.

The first stage of the expanded Star Math development was identifying the set of skills to be assessed. Multiple resources were consulted to determine the set of skills most appropriate for assessing the mathematics development of K–12 US students, typical mathematics curricula, and current mathematics standards. The resources include, but are not limited to:

- *Common Core State Standards for Mathematics*
- National Mathematics Advisory Panel, *Foundations for Success: The final report of the National Mathematics Advisory Panel*
- National Council of Teachers of Mathematics (NCTM), *Curriculum Focal Points for Prekindergarten Through Grade 8 Mathematics*
- NCTM, *Principles and Standards for School Mathematics*
- US State standards from all 50 states, updated annually
- *National Assessment of Educational Progress* (NAEP)
- *Trends in International Mathematics and Science Study* (TIMSS)

The development of the skills list included iterative reviews by mathematicians, mathematics educators, assessment experts, and psychometricians specializing in educational assessment. See "Appendix A: Star Math Blueprint Skills" on page 117 for the Star Math Skills List.

For the purpose of content development, the skills list has been organized into four domains: Numbers and Operations; Algebra; Geometry and Measurement; and Data Analysis, Statistics, and Probability. To ensure appropriate distribution of items within each individual test, the assessment blueprint uses six content domains by treating Numbers, Operations, Geometry, and Measurement as separate domains.

The second development stage included item creation and calibration. Assessment items are developed according to established specifications for grade-level appropriateness and then reviewed to ensure the items meet the specifications. Grade-level appropriateness is determined by multiple factors, including math skill, reading level, cognitive load, vocabulary grade level, sentence structure, sentence length, subject matter, and interest level. All writers and editors have content-area expertise and relevant classroom experience and use those qualifications in determining grade-level appropriateness for subject matter and interest level. A strict development process is maintained to ensure quality item development.

Assessment items, once written, edited, and reviewed, are field tested and calibrated to estimate their Rasch difficulty parameters and goodness of fit to the model.

Following these analyses, each assessment item, along with both traditional and IRT analysis information (including fit plots) and information about the test level, form, and item identifier, are stored in an item statistics database. A panel of content reviewers then examines each item to determine whether the item meets all criteria for use in an operational assessment. More detailed information about the field testing and calibration of Star Math items may be found in the Item and Scale Calibration chapter of this manual.

Star Math and the Reorganization of Objective Clusters

The original version of Star Math organized items into 8 content strands, spanning 17 skill sets and 210 discrete skills. Star Math assesses 790 skills in four standards-based blueprint domains, as outlined inTable 3:

Table 3: Comparison of Domains and Skill Sets: Star Math Progress Monitoring versus Star Math

	Star Math Progress Monitoring Strands	Star Math Blueprint Domains
Skills assessed in:	1. Numeration 2. Computation 3. Word Problems 4. Geometry 5. Measurement 6. Algebra 7. Estimation 8. Data Analysis and Statistics	1. Numbers and Operations 2. Algebra 3. Geometry & Measurements 4. Data Analysis, Statistics & Probability
Skill sets	17	54
Number of skills	210	790

Many of the Star Math Progress Monitoring strands are still represented in the new domains; they are just grouped differently. The organization of Star Math domains and skill sets is modeled after the state standards and the Renaissance Core Progress for Math Learning Progression.

Within each domain, skills are organized into sets of closely related skills sets. The resulting hierarchical structure is blueprint domain, blueprint skill set, and blueprint skill. There are four math domains, 54 skill sets, and 790 skills. See "Appendix A: Star Math Blueprint Skills" on page 117 for a complete list of the Star Math blueprint domains, blueprint skill sets, and blueprint skills.

Calculator and Formula Reference Sheets

For specific Star Math skills, a calculator or formula reference sheet is made available to the student alongside of the test item. Depending on the item and the skill addressed, either the calculator, a formula reference sheet specific to the skill, or both may be used. For the purpose of test validity, these tools are provided in the application rather than the student using their own to ensure that they are used only for appropriately identified skills.

Calculator or Formula Reference sheets are available for two general circumstances: 1) the calculation is overly difficult to perform without either a calculator or a reference chart or 2) the ability to perform the calculations is not the focus of the skill, and the calculations are difficult or time-consuming (e.g., word problems, solving equations, or finding the terms of a sequence).

Formula reference sheets are available for upper-grade skills in which the formula and math relations needed are not expected for student memorization. This decision is based on analysis of the ACT, SAT, ADP, and formula reference sheets used on state end-of-year tests.

An analysis of state assessments produced the following guidelines in determining when a calculator should be made available for Star Math:

Table 4: Determination of Calculator Availability in Star Math

Calculation	Upper Limits of Not Using a Calculator[a]
Division (1–2 step problems)	Divisors may be 1-digit, multiples of 25, fractions with 1-digit denominators, or related to basic math facts (1440/120). Other 2-digit divisors may be included if the division is carried out to only 2 or 3 places.
Multiplication (1–2 step problems)	3-digit by 2-digit, 1-digit by 4-digit (non-zero digits).
Multi-step problems (3+ steps)	2-digit by 2-digit multiplication, 1-digit divisors, other limits listed below.
Powers	2-digit numbers squared, 1-digit numbers raised to the 4th power, 2 or 3 raised to a higher power.
Square roots	Perfect squares related to square of the numbers 1–13 (e.g., square root of 144).
Nth roots	Cube roots resulting in one-digit numbers, nth roots resulting in 2 or 3.
Mean (average)	Up to 6 one- or two-digit numbers or 4 multi-digit numbers.

a. When calculation is not the focus of the skill.

Read-Aloud Audio Guidance

For students challenged by textual reading and the language involved in a Star Math test, read-aloud audio guidance was developed as an accommodation. Read-aloud guidance is turned off for all students by default, but teachers may choose to turn it on either for individual students or an entire class. The accommodation is not intended to be used for all students, blind or low-vision students, but instead is intended to assist teachers to work with students whose language skills are at a lower level than their math skills or who have reading challenges that might prevent them from understanding the item. Audio scripts are not intended to read the entire item aloud for students who cannot read or have extreme visual disabilities.

In order to ensure students receiving read-aloud audio guidance do not have an advantage over other students, some items receive a standard audio prompt of "Choose the best answer." Examples of items receiving this prompt would be if the stem included a single below-grade word such as "solve," or "simplify." Another example would be an item that includes a graphic of a coin and the student is asked to identify the value. Referring to the coin as "a quarter" in the audio prompt may make the item easier for a student who knows a quarter is worth $.25, but cannot identify the quarter visually. For content-specific scripts, only numbers and math expressions embedded within sentences are read. Audio is not included for labels on charts and graphs. Content-specific scripts will be provided for answer choices in items that would pose significant difficulty for struggling readers.

For technical reasons, a single audio file is used for each item requiring audio support, even when audio support contains both the stem and answer options. Students may replay the audio at any time, and may answer the item before the audio has finished playing.

Content Specification: Star Math Progress Monitoring

Item development for the original Star Math Progress Monitoring test predates the bank for Star Math, although both tests were developed with the same overarching goals in mind: to accurately measure the target skill in an accurate and concise manner.

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

The final Star Math content specifications were intended to cover the objectives most frequently found in these four sources. In the end, the Star Math content was organized into eight strands: Numeration Concepts; Computation Processes; Word Problems; Estimation; Data Analysis and Statistics; Geometry; Measurement; and Algebra.

Level of Difficulty: Cognitive Load, Content Differentiation, Depth of Knowledge, and Presentation

Each item is constructed with consideration to cognitive load, content differentiation, and presentation as appropriate for the ability and experience of a typical student at that grade level.

- **Cognitive Load:** Cognitive load involves the type and amount of knowledge and thinking that a student must have and use in order to answer the item correctly. The entire impact of the stem and answer choices must be taken into account.
- **Content Differentiation:** Content differentiation involves the level of detail that a student must address to correctly answer the item. Determining and/or selecting the correct answer should not be dependent on noticing subtle differences in the stem or answer choices.
- **Depth of Knowledge:** Depth of Knowledge is a language system used as an evaluative tool for differentiating among the different levels, 1 through 4, of complexity of specific learning expectations. Items are written to engage students at the targeted depth of knowledge identified for each skill within the assessment.
- **Presentation:** The presentation of the item includes consistent placement of item components, including directions, stimulus components, questions, and answer choices. Each of these should have a typical representation for the discipline area and grade level. The level of visual differentiation needed to read and understand the item components must be grade-level appropriate.

Metadata Requirements and Goals

Due to the restrictions for modifying text, the content may not meet the following goals; however, new item development works to bring the content into alignment with these goals:

- **Gender:** After removing gender-neutral items, an equal number of male and female items should be represented. In addition to names (Sara) and nouns (sisters), gender is also represented by pronoun (she). Gender is not indicated by subject matter or appeal. For instance, an item on cooking is not female unless there is a female character in it.
- **Ethnicity:** The goal is to provide students with an assessment that reflects the ethnic diversity of our school children within the US: 48% White, 15% Black or African American, 27% Hispanic, 5% Middle Eastern, and 5% Asian or Indian. Ethnicity can be based on name or subject matter.

- **Subject:** A variety of subject areas should be present across the items, such as Arts/Humanities, Science, History, Physical Education, Math, and Technology.

Metadata is tagged with codes for Genres, Ethnicity, Occupations, Subjects, Topics, and Regions.

Item Development Guidelines: Star Math

Star Math assesses more than 790 grade-specific blueprint skills. Item development is skill-specific. Each item in the item bank is developed for and clearly aligned to one skill. Answering an item correctly does not require math knowledge beyond the expected knowledge for the skill being assessed. The reading level and math level of the item are grade-level appropriate.The ATOS readability formula is used to identify reading level.

Star Math items are multiple-choice. Most items have four answer choices. An item may have two or three answer choices if appropriate for the skill. Items are distributed among difficulty levels. Correct answer choices are equally distributed by difficulty level.

Item development meets established demographic and contextual goals that are monitored during development to ensure the item bank is demographically and contextually balanced. Goals are established and tracked in the following areas: gender, ethnicity, occupation, age, and disability. Items adhere to strict bias and fairness criteria. Items are free of stereotyping, representing different groups of people in non-stereotypical settings. Items do not refer to inappropriate content that includes, but is not limited to content that presents stereotypes based on ethnicity, gender, culture, economic class, or religion; presents any ethnicity, gender, culture, economic class, or religion unfavorably; introduces inappropriate information, settings, or situations; references illegal activities; references sinister or depressing subjects; references religious activities or holidays based on religious activities; references witchcraft; or references unsafe activities.

The majority of items within a skill are homogeneous in presentation, format, or scenario, but have differing computations. A skill may have two or three scenarios which serve as the basis for homogeneous groupings of items within a skill.
All items for a skill are unique. Text is typically presented as 18-point Arial, but smaller text may be necessary to label charts or graphs. Every complete item is presented on screen with stimulus, stem and answer choices visible. Scroll bars are never used, to minimize cognitive load and confusion created by not having all relevant information available at once. Graphics are included in an item only when necessary to solve the problem.

Item stems meet the following criteria with limited exceptions. When possible, the stem is presented in purely mathematic form or may be limited to a single direction such as “simplify.” When an item requires more complex language, the question is concise, direct, and a complete sentence. The question is written so students can answer it without reading the distractors. Generally, completion (blank) stems are not used. If a completion stem is necessary, the stem contains enough information for the student to complete the stem without reading the distractors, and the completion blank is as close to the end of the stem as possible. The stem does not include verbal or other clues that hint at correct or incorrect distractors. The syntax and grammar are straightforward and appropriate for the grade level.

Negative construction is avoided. The stem does not contain more than one question or part. Concepts and information presented in the items are accurate, up-to-date, and verifiable. This includes but is not limited to dates, measurements, locations, and events.

Distractors meet the following criteria with limited exceptions. All distractors are plausible and reasonable. Distractors do not contain clues that hint at correct or incorrect distractors. Incorrect answers are created based on common student mistakes. Distractors that are not common mistakes may vary between being close to the correct answer or close to a distractor that is the result of a common mistake. Distractors are independent of each other, are approximately the same length, have grammatically parallel structure, and are grammatically consistent with the stem. *None of these, none of the above, not given, all of the above,* and *all of these* are generally avoided as distractors.

Item Development Guidelines: Star Math Progress Monitoring

- When preparing specific items to test student knowledge of the content selected for Star Math Progress Monitoring, several item-writing rules were employed. These rules helped to shape the final appearance of the content and hence became part of the content specifications:
- The first and perhaps most important rule was to have the item content, wording, and format reflect the typical appearance of the content in curricular materials. In some testing applications, one might want the item to look different from how the content typically appears in curricular materials. However, the goal for the Star Math test was to have the items reflect how the content appears in curricular materials that students are likely to have used.
- Second, every effort was made to keep item content simple and to keep the required reading levels low. Although there may be some situations in which one would want to make test items appear complex or use higher levels

of reading difficulty, for the Star Math test, the intent was to simplify when possible.

- Third, efforts were made both in the item-writing and in the item-editing phases to minimize cultural loading, gender stereotyping, and ethnic bias in the items.
- Fourth, the items had to be written in such a way as to be presented in the computer-adaptive format. More specifically, items had to be presentable on the types of computer screens commonly found in schools. This rule had one major implication that influenced item presentation: artwork was limited to fairly simple line drawings, and colors were kept to a minimum.
- Finally, items were all to be presented in a multiple-choice format. Answer choices were to be laid out in either a 4 × 1 matrix, a 2 × 2 matrix, or a 1 × 4 matrix.

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

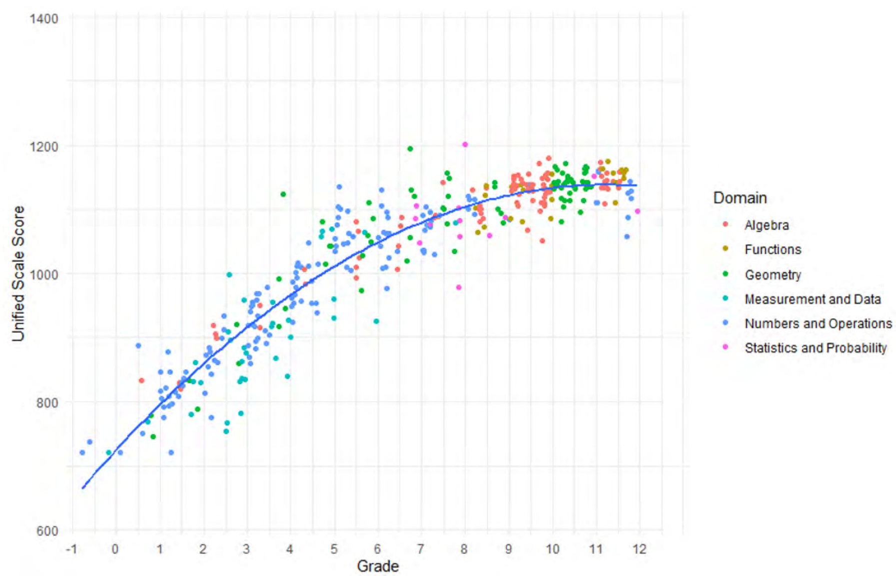
Star Math and Renaissance Learning Progressions for Math

Star Math bridges assessment and instruction through research-based learning progressions to help teachers make effective instructional decisions and to adjust instruction to meet the needs of student at different achievement levels. Star Math assesses more than 790 grade-specific blueprint skills with items developed and aligned to each skill. The skills measured by Star Math are drawn from an overarching pool of skills known as the universal skills pool. The universal skills pool contains the full range of skills reflected in state content standards from all 50 US states and the District of Columbia from early numeracy to high-school level algebra and geometry. The universal skills pool continues to grow and evolve as state standards change and are updated. Learning progressions are created by mapping the skills in the universal skills pool to different content standards. Learning progressions define coherent and continuous pathways in which students acquire knowledge and skills and present the knowledge and skills in teachable orders that can be used to inform instructional decisions.

The first learning progression created for Star Math was the Renaissance Core Progress for Math Learning Progression, which identifies a continuum of math skills that span from early numeracy through high-school level algebra and geometry. It was developed in consultation with leading experts in mathematics by reviewing research and curricular documents and standards, including the National Council of Teachers of Mathematics Curriculum Focal Points, the early work of the National Mathematics Advisory Panel, state and international mathematics standards, and the American Diploma Project Benchmarks. The Renaissance Core Progress for Math Learning Progression is supported by calibration data and psychometric analyses and is regularly refined and updated. Item calibration data from Star Math continually shows that there is a strong correlation between rank ordering of skills in the Renaissance Core Progress for Math Learning Progression and the item difficulty estimates of items written to measure those skills that are used in Star Math.

Figure 1 illustrates the relationship between the sequential order of skills in the Renaissance Core Progress for Math Learning Progression and the average difficulty of the Star Math items measuring that skill on the Star Math Unified scale. Each skill is represented by a single data point with skills in each learning progression domain represented by different color points. The figure shows that skills that are ordered later in the Renaissance Core Progression for Math Learning Progression are often more difficult than skills that are represented earlier in the progression.

Figure 1: Renaissance Core Progress Learning Progression for Math



The relationships shown in Figure 1 continue to evolve as the validation process is ongoing and new Star Math items continue to be written. The continual updating of the Renaissance Core Progress for Math Learning Progression is important to ensure that the ordering of the skills in the Renaissance Core Progress for Math Learning Progression is an accurate representation of the order in which students learn math skills and concepts. To that end, item calibration data collected from Star Math is continuously used to validate and refine learning progressions.

Renaissance now develops individualized learning progressions for all 50 states and the District of Columbia. These state specific learning progressions are also updated yearly as state standards change. The state specific learning progressions cover specific skills represented in each state's grade-level content standards. To create these state specific learning progressions, each state's content standards are analyzed, tagged, and mapped to skills in the universal skills pool. When standards address areas of learning not yet addressed in the universal skills pool, new skills are developed and added to the universal skills pool and potentially added as new Star Math skills. Since Star Math CAT items are written to specific skills which are in turn mapped to skills in the universal skills pool, this allows data from Star Math CAT items to inform state specific learning progression and allows Star Math to report results on state specific content standards and learning progressions. This mapping of Star Math CAT items to skills in the universal skills pool which are in turn mapped to each state's grade-level content standards is one way in which Renaissance works to ensure alignment between Star Math and state content standards.

When a student completes a Star Math assessment, the program uses that student's performance to place the student at the appropriate point in the learning progression designated for that school. This learning progression is usually the state specific learning progression in which the school is located. Locating students in the learning progression helps teachers to identify the skills that students are likely to have already learned and the skills they are ready to learn next. It also indicates whether students are meeting the grade-level performance expectations established by state content standards.

Item and Scale Calibration

Background

Item calibration entails estimating the scaled difficulty of test items by administering them to examinees whose ability is known or estimated, then fitting response models that express the probability of a correct response to each item as a function of examinee ability. To provide accurate item difficulty parameter estimates requires an adequate number of responses to each item, from examinees spanning a broad range of ability. The distribution of ability in the examinee samples need not be closely representative of the distribution of ability in the population, but it needs to be diverse, with large enough numbers of observations above and below the middle of the ability range, as well as from the middle itself.

The introduction of the second generation of Star Math marks the third major evolution in the calibration of Star Math items. For the original 1998 version of Star Math, data for item calibration were collected using printed test booklets and answer sheets, in which the items were formatted to closely match the appearance those items would later take when displayed on computer screens. For the first revision of Star Math in 2002, data collection was done entirely by computer, using a special-purpose application program that administered fixed test forms, but did so on screen, with the same display format and user interface later used in the adaptive version of Star Math 2 (the current Progress Monitoring version). For Star Math versions released since 2011, new test items to be calibrated were embedded as unscored items in Star Math itself, and the data for calibration were collected by the Star Math software. Renaissance Learning calls this data collection process dynamic calibration.

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

In the development of Star Math 2, about 1,100 new items were written. The new items extended the content of the Star Math item bank to include grades 1–12 and expanded the algebra coverage by adding a number of new algebra objectives. Where needed, items measuring other objectives were written to supplement existing items. (Later versions of the program used this same item bank.)

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

For the development of the 34-item Star Math, several thousand new items spanning content appropriate for grades 1–10 were developed. Data for calibrating them were collected using the dynamic calibration feature of Star Math. Using that feature, which was introduced in 2008, small numbers of new, uncalibrated items are randomly selected for each student, and embedded at appropriate random points in Star Math tests. Each student may be administered a small number of these new, uncalibrated items. When a sufficient quantity of response data on the new items has accumulated, calibration analyses take place. Star Math is an application of the Rasch, 1-parameter logistic item response model. For each new item, its location on the Rasch difficulty scale is estimated by fitting a logistic response function to the item responses and Rasch ability scores of the participating examinees. This chapter will describe Rasch item response model, and the criteria applied to screen calibrated items for inclusion in the Star Math item banks. Following that, it will summarize two major item calibration efforts.

The first of these was the calibration of items for use in Star Math Version 2. As noted above, that effort included re-calibration of the original Star Math items, along with new items developed specifically for Star Math 2. Those analyses established the Star Math Rasch ability/item difficulty scale that continues in use today with both versions of Star Math: the 24-item Star Math Progress Monitoring version, an assessment of general math achievement; and the current Star Math, a 34-item standards-based assessment.

The second calibration effort described below was done in advance of the introduction of the current Star Math, a 34-item standards-based version first introduced in 2011. To support the longer test, which assesses a more extensive variety of math skills, a much larger item bank was developed.

The Rasch Item Response Model

In addition to traditional item analyses, the Star Math calibration data are analyzed using item response theory (IRT) methods. Item response theory is widely recognized as the most sophisticated testing approach today.

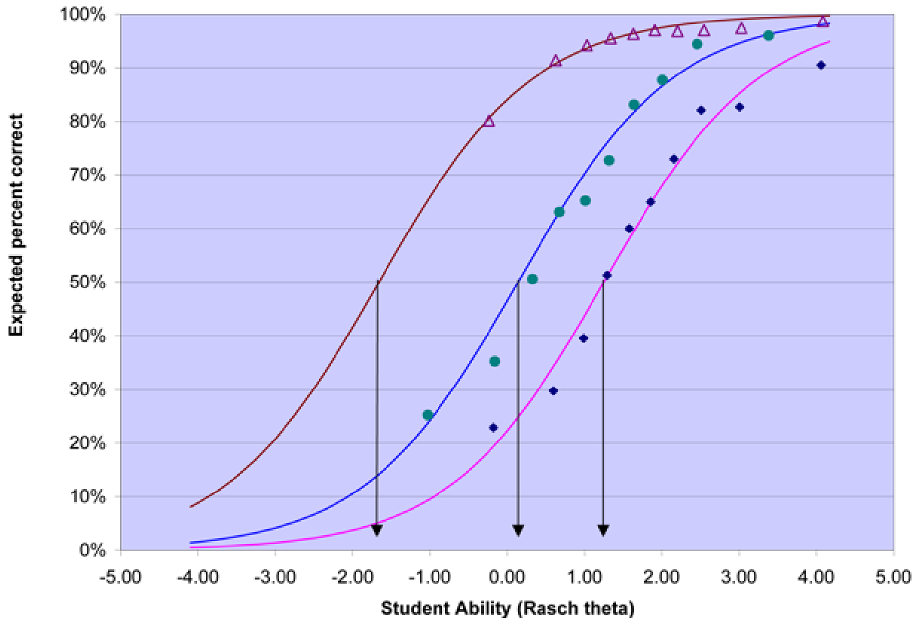
With IRT, the performance of students and the items they answer are placed on the same scale. To accomplish this, every test question is calibrated. Calibration is an IRT-based analytical method for estimating the location of a test question on a common scale used to measure both examinee ability and item difficulty. It is done by administering each question to hundreds and sometimes thousands of students with known performance levels. As a result of calibration, Star "knows" the relative difficulty of every item from kindergarten through grade 12, and expresses it on a developmental scale spanning from the easiest to the hardest questions in the item bank. After taking a Star assessment, a student's score can be plotted on this developmental scale. Placing students and items on the same scale is the breakthrough of IRT because it makes it possible to assign scores on the same scale even though students take different tests. IRT also provides a means to estimate what skills a student knows and doesn't know, without explicitly testing each and every skill.

IRT methods develop mathematical models of the relationship of student ability to the difficulty of specific test questions; more specifically, they model the probability of a correct response to each test question as a function of student ability and item difficulty. Although IRT methods encompass a family of mathematical models, the one-parameter IRT (or Rasch) model was selected for the Star Math data both for its simplicity and its ability to accurately model the performance of the Star Math items.

Within IRT, the probability of answering an item correctly is a function of the student's ability and the difficulty of the item. Since IRT places the item difficulty and student ability on the same scale, this relationship can be represented graphically in the form of an item response function (IRF).

Figure 2 on page 27 is a plot of three item response functions: one for an easy item, one for a more difficult one, and one for an even harder item. Each plot is a continuous S-shaped (ogive) curve. The horizontal axis is the scale of student ability, ranging from very low ability (–5.0 on the scale) to very high ability (+5.0 on the scale). The vertical axis is the percent of students expected to answer each of the three items correctly at any given point on the ability scale. Notice that the expected percent correct increases as student ability increases, but varies from one item to another.

Figure 2: Three Examples of Item Response Functions



In Figure 2, each item's difficulty is the scale point where the expected percent correct is exactly 50. These points are depicted by vertical lines going from the 50% point to the corresponding locations on the ability scale. The easiest item has a difficulty scale value of about –1.67; this means that students located at –1.67 on the ability scale have a 50-50 chance of answering that item right. The scale values of the other two items are approximately +0.20 and +1.25, respectively.

Calibration of test items estimates the IRT difficulty parameter for each test item and places all item parameters onto a single scale used to assess the difficulty of test items, and the ability of students, ranging from Kindergarten through 12th grade level. The difficulty parameter for each item is estimated, along with measures to indicate how well the item conforms to (or "fits") the theoretical expectations of the presumed IRT model.

Also plotted in Figure 2 are the actual percentages of correct responses of groups of students to all three items. Each group is represented as a small triangle, circle, or diamond. Each of those geometric symbols is a plot of the percent correct against the average ability level of the group. Ten groups' data are plotted for each item; the triangular points represent the groups responding to the easiest item. The circles and diamonds, respectively, represent the groups responding to the moderate and to the most difficult item.

Calibration of Items for Star Math Version 2

This section summarizes the psychometric research and development undertaken to prepare the large pool of calibrated math test items for use in Star Math version 2 (now called Star Math Progress Monitoring). As already described above, about 1,100 items spanning grades 1 to 12 were added to the original bank of about 1,400 items and data were collected in the Spring of 2001. The calibration analyses of those items established the underlying Star Math Rasch scale that persists today. The methodology used to develop that scale is summarized below.

Sample Description

To obtain a sample that was representative of the diversity of mathematics achievement in the US school population, school districts, specific schools, and individual students were selected to participate in the Star Math 2 Calibration Study. The sampling frame consisted of all US schools, stratified on three key variables: geographic region of the country, school size, and socioeconomic status. The Star Math calibration sample included students from 261 schools from 45 of the 50 United States. Table 5 and Table 6 present the characteristics of the calibration sample.

Table 5: Sample Characteristics, Star Math 2 Calibration Study—Spring 2001 (N = 44,939 Students)

		Students	
		National %	**Sample %**
Geographic Region	Northeast	20.4%	7.8%
	Midwest	23.5%	22.1%
	Southeast	24.3%	37.3%
	West	31.8%	32.9%
District Socioeconomic Status	Low	28.4%	30.2%
	Average	29.6%	38.9%
	High	31.8%	23.1%
	Non-Public	10.2%	8.1%
School Type and District Enrollment	Public		
	< 200	15.8%	24.2%
	200–499	19.1%	26.2%
	500–1,999	30.2%	26.4%
	2,000 or More	24.7%	15.1%
	Non-Public	10.2%	8.1%

Table 6: Ethnic Group and Gender Participation, Star Math 2 Calibration Study—Spring 2001 (N = 44,939 Students)

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

Item Presentation

The Star Math 2 calibration data were collected by administering test items on screen, with display characteristics identical to those implemented in the earlier Star Math version. However, the calibration items were administered in forms consisting of fixed sequences of items, as opposed to the adaptive testing format.

Seven levels of test forms were constructed corresponding to varying grade levels. Because growth in mathematics is much more rapid in the lower grades, there was only one grade per level for the first four levels. As grade level increases, there is more variation among both students and school curricula, so a single test level can cover more than one grade level. Grades were assigned to test levels after extensive consultation with mathematics instruction experts, and assignments were consistent both with the Star Math item development framework and with assignments used in other math achievement tests. To create the levels of test forms, therefore, items were assigned to grade levels such that resulting test forms sampled an appropriate range of objectives from each of the strands that are typically represented at or near the targeted grade levels. Table 7 on page 30 describes the various test form designations used for the Star Math 2 Calibration Study.

Table 7: Test Form Levels, Grades, Numbers of Items per Form and Numbers of Test Forms, Star Math 2 Calibration Study—Spring 2001

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

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

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

Section A began with items measuring Numeration Concepts and Computation Processes, followed by items measuring Other Applications. Section C reversed this order, with Other Applications items preceding Numeration Concepts and Computation Processes items.

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

With the exception of Levels A and G, approximately half of the vertical anchor items in each form came from the next lower level, and the other half came from the next higher level. Items chosen as vertical anchor items were selected partially based on their difficulty; items expected to be answered correctly by more than 80 percent or fewer than 50 percent of out-of-level students were not used as vertical anchor items. Two versions of each form were used: version A and version B. Each version A form consisted of Sections A, B, and C in that order. Each version B form contained the same items, arranged in reverse order, with Section

C followed by Sections B and A. The alternate forms counterbalanced the order of item presentation, as a defense against possible order effects influencing the psychometric properties of the items. In all three test sections, items were chosen so that content was balanced at each level, with the numbers of items measuring each of the content domains roughly proportional to the distribution of items among the domains at each level.

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

Calibration of New Items for Current Star Math Versions

As described above, beginning in 2008 and continuing with the current version of Star Math, data needed for item calibration have been collected on-line, by embedding small numbers of uncalibrated items within Star Math tests. After sufficient numbers of item responses have accumulated, the Rasch difficulty of each new item is estimated by fitting a logistic model to the item response data and the Star Math Rasch scores of the students' tests. Renaissance Learning calls this overall process "dynamic calibration."

Typically, dynamic calibration is done in batches of several hundred new test items. Each student's test may include between 1 and 5 uncalibrated items. Each item is tagged with a grade level, and is typically administered only to students at that grade level and the next higher grade. The selection of the uncalibrated items to be administered to each student is at random, resulting in nearly equivalent distributions of student ability for each item at a given grade level.

Both traditional and IRT item analyses are conducted of the item response data collected. The traditional analyses yielded proportion correct statistics, as well as biserial and point-biserial correlations between scores on the new items and actual scores on the Star Math tests.

For dynamic calibration, a minimum of 1,000 responses per item is the data collection target. In practice, because of the very large number of Star Math tests administered each year, the average number of students responding to each new test item is typically several times the target. The calibration analysis proceeds one item at a time, using SAS/STAT™ software to estimate the threshold (difficulty) parameter of every new item by calculating the non-linear regression of item scores (0 or 1) on the Star Math Rasch ability estimates. The accuracy of the non-linear regression approach has been corroborated by conducting parallel analyses using Winsteps software. In tests, the two methods yielded virtually identical results.

The "dynamic calibration" approach taken to obtain response data for Star Math new item calibration today is quite different from the approaches taken in the development of item banks for the original Star Math and Star Math 2.

The earlier approaches employed multiple fixed-form field tests as the vehicle for new item response data collection; the analyses themselves fit response models to the new items, using the response data itself as the basis for estimating examinee ability. In today's Star Math, items to be calibrated are embedded as unscored items in Star Math, and the Star Math scores are employed as the ability estimates against which the response models are fit. To ensure a broad diversity of examinee ability, uncalibrated items are selected randomly and administered to students at the target grade level of each item, as well as one grade level above the target, and in some cases one grade level below. Although a nationally representative examinee sample is not required for item calibration, it is useful to evaluate the diversity of the samples who contributed to the calibration data.

This section describes an example of one large dynamic calibration cycle. Table 8, Table 9, and Table 10 on the next page summarize demographic data on about 1.5 million students and 2,473 new items that were part of this process between February 2010 and July 2011. Similar-sized student and item samples were calibrated during other periods, throughout the 2008, 2009, and 2010 school years.

Over 1.5 million students from 7,340 schools in 49 states in addition to Canada and the US Virgin Islands contributed to the overall response data set. Many of those students took two or more Star Math tests during that interval; the total number of tests taken was over 3 million. The number of responses per item ranged from 520 to 58,805, with a median of 2,561.

Of the students participating, 1,446,760 were in US schools; selected demographic data on the U.S. students are in the following tables. Table 8 displays the recorded demographic characteristics of those examinees. Table 9 displays the distribution of the examinees by region of the US; examinees from Canada and outside North America also participated, but their numbers were quite small and are not reported here. Table 10 displays the distribution by gender. Entering the data for each of these analyses was optional; each table tallies only those cases for which the relevant data elements were recorded.

Table 8: Sample Ethnicity, Star Math Calibration Study—February 2010–July 2011 (N = 1,446,760 US Students)

Ethnicity Description	Observations	Observed Percentage	Population Percentage
American Indian or Alaskan Native	16,058	2.99	1.1
Asian or Pacific Islander	16,332	3.04	3.9
Black	156,416	29.13	16.8
Hispanic	105,433	19.64	14.7
Other Race or Ethnicity	1,577	0.29	–
White	241,103	44.90	63.5
Total Observations	536,919		

Table 9: Sample by US Region, Star Math Calibration Study—February 2010–July 2011 (N = 1,446,760 US Students)

Region	Observations	Observed Percentage	Population Percentage
Midwest	169,311	26.13	23.50
Northeast	39,810	6.14	20.40
Southeast	231,819	35.78	24.30
West	207,042	31.95	31.80
Total	647,982		

Table 10: Sample by Gender, Star Math Calibration Study—February 2010–July 2011 (N = 1,446,760 US Students)

Gender	Observations	Observed Percentage	Population Percentage
Female	490,357	48.22	Not available
Male	526,471	51.78	
Total	1,016,828		

Star Math calibration analyses since 2008 followed similar courses. Following extensive quality control checks, the item response data are analyzed using both traditional item analysis techniques and item response theory (IRT) methods. For each test item, the following information is derived using traditional psychometric item analysis techniques:

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

- The correlation between answering the item correctly and the total score (a traditional measure of discrimination).
- The correlation between the endorsement of each alternative answer and the total score.

Traditional Item Difficulty

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

Item Discriminating Power

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

At least two different correlation coefficients are commonly used during item analysis: the point-biserial and the biserial coefficients. The former is a traditional product-moment correlation that is readily calculated, but is known to be somewhat biased in the case of items with p-values that deviate from 0.50. The biserial correlation is derived from the point-biserial and the p-value, and is preferred by many because it in effect corrects for the point-biserial's bias at low and high p-values. For item analysis of Star Math 2 data, the correlation coefficient of choice was the biserial.

Urry (1975) demonstrated that in cases where items could be answered correctly by guessing (e.g., multiple choice items) the value of the biserial correlation is itself attenuated at p-values different from 0.50, and particularly as the p-value approaches the chance level. He derived a correction for this attenuation, which we will refer to as the "Urry biserial correlation." Urry demonstrated that multiple choice adaptive tests are more efficient than conventional tests only if the adaptive tests use items with Urry biserial values that are considerably higher than the target levels often used to select items for conventional test use. His suggestion was to reject items with Urry biserial values lower than 0.62. Item analyses of the Star Math have used the Urry biserial as the correlation coefficient of choice; item selection/rejection decisions have been based in part on his suggested target of 0.62.

Rules for Item Retention

Following these analyses, each test item, along with both traditional and IRT analysis information and information about the test level, form, and item identifier, is stored in a specialized item statistics database system. A panel of internal reviewers then examines each item's statistics to determine whether the item met all criteria for inclusion in the bank of Star Math items. The item statistics database system allows experts easy access to all available information about an item in order to interactively designate items that, in their opinion, meet acceptable standards for inclusion in the Star Math item bank.

Items are eliminated when they meet one or more of the following criteria:

- Item-total correlation (item discrimination) less than the minimum (Urry biserial < 0.62)
- One or more incorrect answer options has a positive item discrimination value
- Sample size of students responding to the item less than 1,000
- The traditional item difficulty indicates that the item is too difficult or too easy
- The item does not appear to fit the Rasch IRT model

In the case of the batch of 2,473 items used in the example of Star Math item calibration above, 884 items (36%) met all the retention rules above, and were accepted for operational use as part of the Star Math adaptive test item bank. Another 538 items met all criteria except the Urry biserial target. Such items would meet commonly applied criteria for use in most conventional tests; those 538 items were retained for use for certain analytical purposes, but will not be used for adaptive testing in Star Math.

Computer-Adaptive Test Design

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

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

During an adaptive test, a student may be "routed" to items at the lowest level of difficulty within the overall pool of items, dependent upon the student's unfolding performance during the testing session. In general, when an item is answered correctly, the student is routed to a more difficult item. When an item is answered incorrectly, the student is instead routed to an easier item. In the case of Star Math, the brancher selects items with a 67 percent expectation of a correct response, given the student's estimated ability, and the item's calibrated difficulty.

A Star Math test consists of a fixed-length, 34-item adaptive test; Star Math Progress Monitoring tests are 24 items in length. Students who have not taken a Star Math test within 180 days initially receive an item whose difficulty level is relatively easy for students at that grade level. This minimizes any effects of initial anxiety that students may have when starting the test and serves to better facilitate the students' initial reactions to the test. The starting points vary by grade level and are based on research conducted as part of the norming process.

When a student has taken a Star Math test within the previous 180 days, the appropriate starting point is based on his or her previous test score information. Following the administration of the initial item, and after the student has entered an answer, the program determines an updated estimate of the student's math achievement level. Then, it selects the next item randomly from among all of the available items having a difficulty level that closely match a target based on the estimated achievement level. Randomization of items with difficulty values near the target level allows the program to avoid overexposure of test items.

Items that have been administered to the same student within the past 120 days are not available for administration. In addition, to avoid frustration, items that are intended to measure advanced mathematical concepts and operations that are more than three grade levels beyond the student's grade level, as determined by where such concepts or operations are typically introduced in math textbooks, are also not available for administration. Because the item pools make a large number of items available for selection, these minor constraints have a negligible impact on the quality of each Star Math computer-adaptive test.

Scoring in the Star Math Tests

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

This approach to scoring enables the software to provide Scaled Scores that are statistically consistent and efficient. Scaled Scores are expressed on a common scale that spans all grade levels covered by the Star Math test. Because the software expresses Scaled Scores on a common scale, Scaled Scores are directly comparable with each other, regardless of grade level. Other scores, such as Percentile Ranks and Grade Equivalents, are derived from the Scaled Scores obtained during the Star Math norming studies.

A New Scale for Reporting Star Math Test Scores

In 1998, Renaissance Learning released the initial 24-item version of Star Math. In 2011, the 34-item standards-based Star Math test was published. Although Star Math measures constructs that are different from those assessed in Star Reading, a common scale—the Unified Score Scale—that can be used to report scores on both tests was recently developed. The Unified Score Scale was introduced into use in the 2017–2018 school year as an optional alternative scale for reporting achievement on all Star tests.

The Unified Score Scale is derived from the Star Reading Rasch scale of ability and difficulty, which was first introduced with the development of Star Reading Version 2.

The Unified Star Math scale was developed by performing the following steps:

- The Rasch scale used by Star Math was linked (transformed) to the Star Reading Rasch scale.

- A linear transformation of the transformed Rasch scale was developed that spans the entire range of knowledge and skills measured by both Star Math and Star Reading.

Details of these two steps are presented below.

1. The Rasch scale used by Star Math was linked to the Star Reading Rasch scale.

   In this step, a linear transformation of the Star Math Rasch scale to the Rasch scale used by Star Reading was developed, using a method for linear equating of IRT (item response theory) scales described by Kolen and Brennan (2004, pages 162–165).

2. Because Rasch scores are expressed as decimal fractions, and may be either negative or positive, a more user-friendly scale score was developed that uses positive integer numbers only. A linear transformation of the extended Star Reading Rasch scale was developed that spans the entire range of knowledge and skills measured by both Star Math and Star Reading. The transformation formula is as follows:

   Unified Scale Score = INT (42.93 * Star Reading Rasch Score + 958.74)

   where the Star Reading Rasch score has been extended downwards to values as low as –20.00.

   Following are some features and considerations in the development of that scale, called here the "Unified scale."

   a. For both Star Math and Star Reading, the range of reported Unified scales is from 600 to 1400. Anchor points were chosen such that the Unified scale score of 600 is approximately equivalent to a Star Math scale score of 0, and a Unified score of 1400 is the approximate equivalent of 1300 on the Star Math scale.

   b. The scale is extensible upwards and downwards. Currently, the highest reported Star Math Unified scale score is 1400, but there is no theoretical limit: if Star Math content were extended beyond the high school level, the range of the new scale can be extended upward without limit, as needed. The lowest point is now set at 600, but the Unified scale can readily be extended downward as low as 0, if a reason arises to do so.

Table 11 contains a table of selected Star Math Rasch ability scores and their equivalents on the Star Math and Unified Score scales.

Table 11: Some Star Math Rasch Scores and their Equivalents on the Star Math and Unified Score Scales

Minimum Rasch Score	Star Math Scaled Score	Unified Scale Score
–8.35	0	600
–7.72	50	638
–7.08	100	668
–6.45	150	699
–5.81	200	730
–5.18	250	761
–4.54	300	791
–3.91	350	822
–3.27	400	853
–2.64	450	884
–2.00	500	914
–1.37	550	945
–0.74	600	976
–0.10	650	1007
0.54	700	1037
1.17	750	1068
1.81	800	1099
2.44	850	1130
3.07	900	1160
3.71	950	1191
4.34	1000	1222
4.98	1050	1253
5.61	1100	1283
6.25	1150	1314
6.88	1200	1345
7.52	1250	1376
8.15	1300	1400

Reliability and Measurement Precision

Measurement is subject to error. A measurement that is subject to a great deal of error is said to be imprecise; a measurement that is subject to relatively little error is said to be *reliable*. In psychometrics, the term *reliability* refers to the degree of measurement precision, expressed as a proportion. A test with perfect score precision would have a reliability coefficient equal to 1, meaning that 100 percent of the variation among persons' scores is attributable to variation in the attribute the test measures, and none of the variation is attributable to error. Perfect reliability is probably unattainable in educational measurement; for example, a test with a reliability coefficient of 0.90 is more likely. On such a test, 90 percent of the variation among students' scores is attributable to the attribute being measured, and 10 percent is attributable to errors of measurement. Another way to think of score reliability is as a measure of the consistency of test scores. Two kinds of consistency are of concern when evaluating a test's measurement precision: internal consistency and consistency between different measurements. First, internal consistency refers to the degree of confidence one can have in the precision of scores from a single measurement. If the test's internal consistency is 95 percent, just 5 percent of the variation of test scores is attributable to measurement error.

Second, reliability as a measure of consistency between two different measurements indicates the extent to which a test yields consistent results from one administration to another and from one test form to another. Tests must yield somewhat consistent results in order to be useful; this reliability coefficient is obtained by calculating the coefficient of correlation between students' scores on two different occasions, or on two alternate versions of the test given at the same occasion.

Because the amount of the attribute being measured may change over time, and the content of tests may differ from one version to another, the internal consistency reliability coefficient is generally higher than the correlation between scores obtained on different administrations.

There are a variety of methods of estimating the reliability coefficient of a test. Methods such as Cronbach's alpha and split-half reliability are single administration methods and assess internal consistency. Coefficients of correlation calculated between scores on alternate forms, or on similar tests administered two or more times on different occasions, are used to assess alternate forms reliability, or test-retest reliability (stability).

In a computerized adaptive test such as Star Math, content varies from one administration to another, and it also varies with each student's performance. Another feature of computerized adaptive tests based on Item Response Theory

(IRT) is that the degree of measurement error can be expressed for each student's test individually.

The Star Math tests provide two ways to evaluate the reliability of scores: reliability coefficients, which indicate the overall precision of a set of test scores, and conditional standard errors of measurement (CSEM), which provide an index of the degree of error in an individual test score. A reliability coefficient is a summary statistic that reflects the average amount of measurement precision in a specific examinee group or in a population as a whole. In Star Math, the CSEM is an estimate of the unreliability of each individual test score. While a reliability coefficient is a single value that applies to the test in general, the magnitude of the CSEM may vary substantially from one person's test score to another's. Another part of evaluating reliability is looking at the reliability of classification decisions. In many applications of Star Math, three normative benchmarks, set at the 10th, 25th, and 40th percentile ranks, are used to classify students into the performance categories of intensive intervention, intervention, on watch, and at/above benchmark. These classifications are often used in a response-to-intervention (RTI) and multi-tiered system of supports (MTSS) framework by schools. To look at reliability of classifications based on benchmarks, decision accuracy and decision consistency indices can be computed. Like reliability coefficients based on test scores, decision accuracy and consistency indices range from 0 to 1 with values close to 1 indicating more accurate and consistent classifications.

This chapter presents three different types of reliability coefficients: generic reliability, split-half reliability, and alternate form reliability. This is followed by statistics on the conditional standard error of measurement of Star Math test scores. The chapter also presents indices of decision accuracy and consistency.

The reliability and measurement error presentation is divided into two sections below: First is a section describing the reliability coefficients, standard errors of measurement, and decision accuracy and consistency indices for the 34-item Star Math tests. Second, another brief section presents reliability coefficients, standard errors of measurement, and decision accuracy and consistency indices for the 24-item Star Math progress monitoring tests.

34-Item Star Math Tests

Generic Reliability

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

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

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

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

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

Using this technique with the Star Math 2018–2019 school year data resulted in the generic reliability estimates shown in the third column of Table 12. Because this method is not susceptible to error variance introduced by repeated testing, multiple occasions, and alternate forms, the resulting estimates of reliability are generally higher than the more conservative alternate form reliability coefficients. These generic reliability coefficients are, therefore, plausible upper-bound estimates of the internal consistency of the Star Math computerized adaptive test.

Table 12: Reliability Estimates from the Star Math 2018–2019 Data on both the Unified Scale and the Enterprise Scale

Grade	Reliability Estimates: For Both Unified and Enterprise Scales						
	Generic		Split-Half		Alternate Form		
	N	ρxx	N	ρxx	N	ρxx	Average Days between Testing
K	50,000	0.90	20,000	0.90	10,000	0.71	102
1	1,000,000	0.89	20,000	0.89	172,500	0.75	97
2	1,000,000	0.90	20,000	0.90	250,000	0.80	92
3	1,000,000	0.91	20,000	0.91	250,000	0.82	90
4	1,000000	0.91	20,000	0.91	250,000	0.83	91
5	1,000,000	0.92	20,000	0.92	250,000	0.85	92
6	1,000,000	0.92	20,000	0.92	225,000	0.85	101
7	1,000,000	0.93	20,000	0.93	200,000	0.85	106
8	1,000,000	0.93	20,000	0.93	200,000	0.85	105
9	400,000	0.93	20,000	0.93	85,000	0.85	113
10	400,000	0.93	20,000	0.93	65,000	0.85	114
11	100,000	0.93	20,000	0.93	40,000	0.84	113
12	100,000	0.94	20,000	0.94	20,000	0.83	109
Overall	9,050,000	0.98	260,000	0.98	2,017,500	0.94	98

Generic reliability estimates for scores on both the Unified score scale and the Enterprise score scale are shown in Table 12. Because both the Unified scaled and the Enterprise scale are linear transformations of the underlying Star Math Rasch scores, the reliability estimates are the same across both scales. Results in Table 12 indicate that the overall generic reliability of the scores was about 0.98. Coefficients ranged from a low of 0.89 in grade 1 to a high of 0.94 in grade 12.

As the data in Table 12 show, Star Math generic reliability is quite high, grade by grade and overall. Star Math also demonstrates high consistency between alternate forms as shown in the rightmost columns of the same table. Star Math's technical quality for an interim assessment is on a virtually equal footing with the highest-quality summative assessments in use today.

Split-Half Reliability

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

A split-half reliability coefficient is calculated in three steps. First, the test is divided into two halves, and scores are calculated for each half. Second, the correlation between the two resulting sets of scores is calculated; this correlation is an estimate of the reliability of a half-length test. Third, the resulting reliability value is adjusted, using the Spearman-Brown formula, to estimate the reliability of the full-length test.

In internal simulation studies, the split-half method provided accurate estimates of the internal consistency reliability of adaptive tests, and so it has been used to provide estimates of Star Math reliability. These split-half reliability coefficients are independent of the generic reliability approach discussed earlier and more firmly grounded in the item response data. Split-half scores were based on all of the 34 items of the Star Math tests; scores based on the odd- and the even-numbered items were calculated separately. The correlations between the two sets of scores were corrected to a length of 34 items, yielding the split-half reliability estimates displayed in Table 12.

Results indicated that the overall split-half reliability of scores was 0.98. The coefficients ranged from a low of 0.89 in grade 1 to a high of 0.94 in grade 12. These reliability estimates were consistently high across grades 1–12, as a result of the measurement efficiency inherent in the adaptive nature of the Star Math test.

Alternate Form Reliability

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

Content sampling, temporal changes in individuals' performance, and growth or decline over time can affect alternate form reliability coefficients, usually making them appreciably lower than internal consistency reliability coefficients. The alternate form reliability study provided estimates of Star Math reliability using a variation of the test-retest method. In the traditional approach to test-retest reliability, students take the same test twice, within a short time interval, usually a few days, between administrations. In contrast, the Star Math alternate form reliability study administered two different tests by avoiding during the second test the use of any items the student had encountered in the first test. All other aspects of the two tests were identical. The correlation coefficient between the scores on the two tests was taken as the reliability estimate.

The alternate form reliability estimates for the Star Math test were calculated using the Star Math Unified scale scores. Checks were made for valid test data on both test administrations and cases of apparent motivational discrepancies were removed.

Table 12 on page 43 includes overall and within-grade alternate form reliability, along with an indication of the average number of days between testing occasions, ranging from 91–114 days. Results indicated that the overall reliability of the scores was about 0.94. The alternate form coefficients ranged from a low of 0.71 in grade K to a high of 0.85 in grades 5 to 10.

Because errors of measurement due to content sampling and temporal changes in individuals' performance can affect this correlation coefficient, this type of reliability estimate provides a conservative estimate of the reliability of a single Star Math administration. In other words, the actual Star Math reliability is likely higher than what the alternate form reliability estimates indicate.

Star Math was designed to be a standards-based assessment, meaning that its item bank measures skills identified by exhaustive analysis of national and state standards in Math, from grades K–12, including Algebra and Geometry. The 34-item Star Math content covers many more skills than previous versions of Star Math, which administered only 24 items.

The increased length of the current version of Star Math, combined with its increased breadth of skills coverage and enhanced technical quality, was expected to result in improved measurement precision; this showed up as slightly increased reliability, in both internal consistency reliability and alternate form reliability as shown in Table 12. For comparison, see Table 16.

Standard Error of Measurement

When interpreting the results of any test instrument, it is important to remember that the scores represent estimates of a student's true ability level. Test scores are not absolute or exact measures of performance. Nor is a single test score infallible in the information that it provides. The standard error of measurement can be thought of as a measure of how precise a given score is. The standard error of measurement describes the extent to which scores would be expected to fluctuate because of chance. If measurement errors follow a normal distribution, an SEM of 18 means that if a student were tested repeatedly, his or her scores would fluctuate within 18 points of his or her true score about 68 percent of the time, and within 36 points (twice the SEM) roughly 95 percent of the time. Since reliability can also be regarded as a measure of precision, there is a direct relationship between the reliability of a test and the standard error of measurement for the scores it produces: as reliability increases, standard error of measurement decreases.

The Star Math tests differ from traditional tests in at least two respects with regard to the standard error of measurement. First, Star Math software computes the SEM for each individual student based on his or her performance, unlike most traditional fixed tests that report the same SEM value for every examinee. Each administration of Star Math yields a unique "conditional" SEM (CSEM) that reflects the amount of information estimated to be in the specific combination of items that a student received in his or her individual test. Second, because the Star Math test is adaptive, the CSEM will tend to be lower than that of a conventional test, particularly at the highest and lowest score levels, where conventional tests' measurement precision is weakest. Because the adaptive testing process attempts to provide equally precise measurement, regardless of the student's ability level, the CSEM of scores are very similar for all students.

Table 13 and Table 14 summarize the distribution of average CSEM values for the 2018–2019 data. The average CSEM on the Unified scale was 18 scaled score units overall and quite consistent across grades, ranging from a low of 18 in grades K–10 to a high of 19 in grades 11 and 12 (Table 13). The average CSEM on the Enterprise scale was 30 scaled score units for every grade (Table 14).

Alternatively, Table 13 and Table 14 also report global SEM values, which were computed using the traditional SEM estimation method based on internal consistency reliability and the variance of test scores as follows:

$$SEM = SQRT(1 - \rho)\ \sigma x$$

where

SQRT() is the square root operator

ρ is the estimated internal consistency reliability

σx is the standard deviation of the observed scores (in this case, Scaled Scores)

The global SEMs were almost identical to the average CSEMs on both the Unified and Enterprise scales.

Because the standard error of measurement (SEM) is scale dependent, there are differences in the reported SEMs between the Star Math Unified and Enterprise scales. Overall, the lower SEM values in Table 13 compared to those in Table 14 reflect the differences between the Unified and Enterprise scale score ranges. Neither of these is "better," as the reliability estimates are the same for both scales.

Table 13: Standard Error of Measurement for the 2018–2019 Star Math Data on the Unified Scale

Grade	Sample Size	Standard Error of Measurement—Unified Scale		
		Conditional		Global
		Average	Standard Deviation	
K	50,000	18	1.3	18
1	1,000,000	18	1.1	18
2	1,000,000	18	1.2	18
3	1,000,000	18	1.2	18
4	1,000,000	18	1.3	18
5	1,000,000	18	1.4	18
6	1,000,000	18	1.3	18
7	1,000,000	18	1.5	18
8	1,000,000	18	1.6	18
9	400,000	18	1.4	18
10	400,000	18	1.6	18
11	100,000	19	1.7	18
12	100,000	19	1.7	19
All	9,050,000	18	1.4	16

Table 14: Standard Error of Measurement for the 2018–2019 Star Math Data on the Enterprise Scale

Grade	Sample Size	Standard Error of Measurement—Enterprise Scale		
		Conditional		Global
		Average	Standard Deviation	
K	50,000	30	2.1	30
1	1,000,000	30	1.9	30
2	1,000,000	30	2.0	30
3	1,000,000	30	1.9	29
4	1,000,000	30	2.0	30
5	1,000,000	30	2.2	30
6	1,000,000	30	2.2	30
7	1,000,000	30	2.5	30
8	1,000,000	30	2.6	30
9	400,000	30	2.3	30
10	400,000	30	2.6	30
11	100,000	30	2.8	30
12	100,000	30	2.8	31
All	9,050,000	30	2.2	26

Decision Accuracy and Decision Consistency

Decision accuracy is generally defined as the degree to which observed examinee classification decisions on a single assessment would agree with true classifications for a given set of cut scores. There are multiple approaches to estimate decision accuracy. Star Math uses Rudner's index (Rudner, 2001; 2005) based on item response theory (IRT), which assumes that the maximum likelihood estimate of ability converges to a normal distribution with mean equal to θ and standard deviation equal to the conditional standard error of measurement (CSEM). Mathematically, this index can be computed as:

$$\hat{\tau} = \sum (\hat{\mathbf{P}} * \mathbf{W})/N_e,$$

where Σ denotes the summation of all matrix elements, * denotes element-wise matrix multiplication, Ne is the number of examinees, $\hat{\mathbf{P}}$ is a $Ne \times C$ matrix of expected probabilities with C being the number of performance categories on the assessment, and **W** is a $Ne \times C$ matrix of binary weights used to indicate the observed performance categories on the assessment. The $\hat{\mathbf{P}}$ matrix is defined as:

$$\hat{\mathbf{P}} = \begin{bmatrix} \hat{p}11 & \hat{p}12 & \cdots & \hat{p}1C \\ \hat{p}21 & \hat{p}22 & \cdots & \hat{p}2C \\ \vdots & \vdots & \cdots & \vdots \\ \hat{p}N_e1 & \hat{p}N_e2 & \cdots & \hat{p}N_eC \end{bmatrix}$$

with the expected probability $\hat{p}ic$ in the above matrix estimated as:

$$\hat{p}ic = \phi(\kappa ic, \kappa i(c+1), \hat{\theta}i, \hat{\sigma}\theta_i),$$

where $\phi(a, b, \mu, \sigma)$ is the area from a to b under a normal curve with a mean of μ and a standard deviation of σ, $\hat{\theta}i$ is examinee i's IRT ability estimate, $\hat{\sigma}\theta_i$ is the corresponding CSEM for the ability estimate $\hat{\theta}i$, and κic and $\kappa i(c+1)$ are cut scores with $\kappa i1 = -\infty$, $\kappa i2$ being the cut score separating performance categories 1 and 2, $\kappa i3$ being the cut score separating performance categories 2 and 3, and so on with the last cut score $\kappa i(c+1) = \infty$. The **W** matrix of weights is defined as:

$$\mathbf{W} = \begin{bmatrix} w11 & w12 & \cdots & w1C \\ w21 & w22 & \cdots & w2C \\ \vdots & \vdots & \cdots & \vdots \\ wN_e1 & wN_e2 & \cdots & wN_eC \end{bmatrix},$$

where the weight, *wic*, equals 1 if the student was classified into performance level category *C* based on their ability estimate and 0 otherwise.

A counterpart to decision accuracy is decision consistency, defined as the degree to which examinees would be classified into the same performance categories given parallel replications of the same assessment. The method used to estimate decision consistency is based on an extension to Rudner's decision accuracy index, which is described in Wyse and Hao (2012). This index can be estimated as:

$$\hat{\gamma} = \sum (\hat{\mathbf{P}} * \hat{\mathbf{P}})/N_e,$$

where *Ne* is the number of examinees and $\hat{\mathbf{P}}$ is the same *Ne* × *C* matrix of expected probabilities used when computing the decision accuracy index.

For Star Math, three different classification decisions based on benchmarks set at the 10th, 25th, and 40th percentile ranks in the student norms are available by default in the Star Math software. These cut scores are used to separate students into four different performance categories: intensive intervention, intervention, on watch, and at/above benchmark. Table 15 shows estimates of decision accuracy and consistency when identifying students based on the three individual benchmarks as well as all three benchmarks together using random samples of students that took Star Math in the 2018–2019 school year.

Results indicate that decision accuracy and consistency were quite high overall and across grades. For PR10, decision accuracy was 0.97 for every grade, while decision consistency ranged from 0.95 to 0.96. For PR25, decision accuracy ranged from a low of 0.93 to a high of 0.95, while decision consistency ranged from 0.90 to 0.93. For PR40, decision accuracy ranged from a low of 0.90 to a high of 0.93, while decision consistency ranged from 0.86 to 0.90. Decision accuracy when using all three benchmarks together ranged from a low of 0.81 to a high of 0.85, while decision consistency ranged from a low of 0.74 to a high of 0.79. These are high levels of decision accuracy and consistency when making classification decisions based on each individual benchmark or all three benchmarks together, and support using Star Math in RTI/MTSS frameworks.

Table 15: Decision Accuracy and Consistency for Different Benchmarks Based on 2018–2019 Star Math Tests

Grade	N	Decision Accuracy				Decision Consistency			
		PR10	PR25	PR40	All 3 Benchmarks	PR10	PR25	PR40	All 3 Benchmarks
K	50,000	0.97	0.93	0.91	0.83	0.95	0.91	0.88	0.79
1	1,000,000	0.97	0.94	0.91	0.83	0.96	0.91	0.87	0.78
2	1,000,000	0.97	0.93	0.90	0.81	0.95	0.90	0.86	0.75
3	1,000,000	0.97	0.93	0.91	0.82	0.96	0.90	0.87	0.76
4	1,000,000	0.97	0.93	0.91	0.82	0.96	0.90	0.87	0.76
5	1,000,000	0.97	0.93	0.91	0.82	0.95	0.91	0.88	0.76
6	1,000,000	0.97	0.94	0.92	0.83	0.96	0.91	0.88	0.77
7	1,000,000	0.97	0.94	0.92	0.83	0.96	0.92	0.89	0.78
8	1,000,000	0.97	0.94	0.92	0.83	0.96	0.92	0.89	0.78
9	400,000	0.97	0.95	0.93	0.84	0.96	0.93	0.90	0.78
10	400,000	0.97	0.95	0.92	0.84	0.96	0.93	0.89	0.78
11	100,000	0.97	0.95	0.92	0.85	0.96	0.93	0.89	0.79
12	100,000	0.97	0.95	0.92	0.85	0.96	0.93	0.89	0.79
Overall	9,050,000	0.97	0.94	0.92	0.83	0.96	0.92	0.88	0.77

24-Item Star Math Progress Monitoring Tests

Star Math is used for both universal screening and progress monitoring. The 34-item Star Math test is widely used for universal screening. A shorter version—the 24-item Star Math progress monitoring test—exists for use in progress monitoring. The following section summarizes the reliability and the standard error of measurement of the progress monitoring version of Star Math.

Reliability Coefficients

Table 16 shows the reliability estimates of the Star Math progress monitoring tests from 2017 to 2019 on both the Unified scale and the Enterprise scale.

Table 16: Reliability Estimates from the Star Math Progress Monitoring 2017–2018 and 2018–2019 Data on both the Unified Scale and the Enterprise Scale

Grade	Progress Monitoring Reliability Estimates for Both the Unified and Enterprise Scale			
	Generic		Split-Half	
	N	ρxx	N	ρxx
1	9,000	0.84	8,600	0.86
2	17,000	0.79	16,800	0.81
3	16,000	0.80	15,000	0.81
4	18,000	0.82	17,100	0.83
5	15,000	0.82	14,700	0.82
6	10,000	0.83	9,700	0.83
7	8,500	0.87	8,200	0.87
8	7,500	0.87	7,300	0.87
9	875	0.89	835	0.90
10	730	0.90	662	0.90
11	500	0.91	479	0.92
12	360	0.83	344	0.83
Overall	103,465	0.93	99,720	0.93

The progress monitoring Star Math reliability estimates are also quite high and consistent across grades 1–12, for a test composed of only 24 items.

Overall, these coefficients also compare very favorably with the reliability estimates provided for other published math achievement tests, which typically contain far more items than the 24-item Star Math progress monitoring tests. The Star Math progress monitoring test's high reliability with minimal testing time is a result of careful test item construction and an effective and efficient adaptive-branching procedure.

Standard Error of Measurement

Table 17 and Table 18 show the conditional standard error of measurement (CSEM) and the global standard error of measurement (SEM), overall and by grade level.

Table 17: Standard Error of Measurement for the 2017–2018 and 2018–2019 Star Math Progress Monitoring Data on the Unified Scale

Grade	Progress Monitoring Standard Error of Measurement—Unified Scale				
	Conditional			Global	
	Sample Size	Average	Standard Deviation	Sample Size	SEM
1	9,000	21	2.0	8,600	19
2	17,000	21	2.0	16,800	20
3	16,000	21	2.1	15,000	20
4	18,000	21	2.2	17,100	20
5	15,000	21	2.4	14,700	21
6	10,000	21	2.6	9,700	21
7	8,500	21	2.7	8,200	21
8	7,500	21	2.7	7,300	21
9	875	21	2.8	835	20
10	730	21	2.9	662	21
11	500	22	3.1	479	20
12	360	21	2.6	344	21
All	103,465	21	2.3	99,720	20

Table 18: Standard Error of Measurement for the 2017–2018 and 2018–2019 Star Math Progress Monitoring Data on the Enterprise Scale

Grade	Progress Monitoring Standard Error of Measurement—Enterprise Scale				
	Conditional			Global	
	Sample Size	Average	StandardDeviation	Sample Size	SEM
1	9,000	38	3.6	8,600	36
2	17,000	38	3.7	16,800	36
3	16,000	38	3.8	15,000	37
4	18,000	39	4.1	17,100	37
5	15,000	39	4.4	14,700	39
6	10,000	39	4.7	9,700	39
7	8,500	39	4.9	8,200	39
8	7,500	39	4.9	7,300	39
9	875	39	5.2	835	37
10	730	39	5.3	662	39
11	500	40	5.6	479	38
12	360	39	4.8	344	39
All	103,465	39	4.3	99,720	37

Comparing the estimates of reliability and measurement error of Star Math (Table 12, Table 13, and Table 14) with those of Star Math progress monitoring (Table 16, Table 17, and Table 18) confirms that Star Math is slightly superior to the shorter Star Math progress monitoring assessments in terms of reliability and measurement precision.

Decision Accuracy and Consistency

Table 19 shows the decision accuracy and consistency indices for PR10, PR25, and PR40 benchmarks for Star Math Progress Monitoring based on data collected in the 2017–2018 and 2018–2019 school years. Results suggest that the decision accuracy and consistency for the Star Math Progress Monitoring tests were high, but slightly lower than the values observed for the 34-item Star Math tests. These high levels of decision accuracy and consistency support using Star Math tests in RTI/MTSS frameworks.

Table 19: Decision Accuracy and Consistency for Different Benchmarks Based on 2017–2018 and 2018–2019 Star Math Progress Monitoring Tests

Grade	N	Decision Accuracy				Decision Consistency			
		PR10	PR25	PR40	All 3 Benchmarks	PR10	PR25	PR40	All 3 Benchmarks
1	9,000	0.97	0.93	0.90	0.81	0.95	0.90	0.86	0.76
2	17,000	0.96	0.91	0.87	0.77	0.95	0.87	0.82	0.70
3	16,000	0.94	0.88	0.86	0.71	0.92	0.83	0.81	0.63
4	18,000	0.94	0.89	0.87	0.72	0.92	0.84	0.81	0.64
5	15,000	0.95	0.88	0.87	0.72	0.93	0.83	0.82	0.64
6	10,000	0.94	0.87	0.87	0.70	0.91	0.82	0.82	0.61
7	8,500	0.94	0.88	0.89	0.72	0.91	0.84	0.84	0.64
8	7,500	0.93	0.88	0.90	0.72	0.90	0.84	0.85	0.64
9	875	0.92	0.89	0.90	0.72	0.89	0.84	0.85	0.63
10	730	0.91	0.91	0.92	0.74	0.88	0.87	0.88	0.65
11	500	0.92	0.90	0.91	0.74	0.89	0.86	0.87	0.65
12	360	0.94	0.87	0.90	0.71	0.92	0.81	0.86	0.61
Overall	103,465	0.94	0.89	0.89	0.73	0.91	0.85	0.84	0.65

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.

Content Validity

One touchstone is content validity, which is the relevance of the test questions to the attributes or dimensions intended to be measured by the test. The content of the item bank and the content balancing specifications that govern the administration of each test together form the foundation for “content validity” for the Star Math assessments. These content validity issues were discussed in detail in “Content and Item Development” and were an integral part of the test items that are the basis of the current Star Math version.

Construct Validity

Construct validity, which is the overarching criterion for evaluating a test, investigates the extent to which a test measures the construct(s) that it claims to be assessing. Establishing construct validity involves the use of data and other information external to the test instrument itself. For example, Star Math claims to provide an estimate of a child’s mathematics achievement level. Therefore, demonstration of Star Math construct validity rests on the evidence that the test provides such estimates. There are a number of ways to demonstrate this.

Since mathematics ability varies significantly within and across grade levels and improves as a student’s grade placement increases, scores within Star Math should demonstrate these anticipated internal relationships; in fact, they do. Additionally, scores for Star Math should correlate highly with other accepted measures of mathematics achievement and competence. This section deals with both internal and external evidence of the validity of Star Math as an assessment of Mathematics achievement and competence.

Internal Evidence: Evaluation of Unidimensionality of Star Math

Star Math is a 34-item computerized-adaptive assessment that measures mathematics achievement. Its items are selected adaptively for each student, from a very large bank of mathematics test items, each of which is aligned to one of four blueprint domains:

- Numeration & Operations (NUM)
- Algebra (ALG)
- Geometry & Measurement (GEO)
- Data Analysis, Statistics & Probability (DAT)

Star Math is an application of item response theory (IRT); each test item's difficulty has been calibrated using the unidimensional Rasch model. Therefore, an important assumption is that a test measures only a single construct, specifically mathematics achievement in the case of Star Math. This assumption was tested using factor analysis, which is a statistical technique that can be used to determine the number of dimensions or constructs that a test measures. Both exploratory and confirmatory factor analyses were conducted across grade bands K to 2, 3 to 5, 6 to 8, and 9 to 12.

To begin, a large sample of student Star Math data was assembled. The overall sample consisted of 202,000 student records, which was investigated with confirmatory factor analysis to determine unidimensionality followed by a variety of exploratory factor analyses.

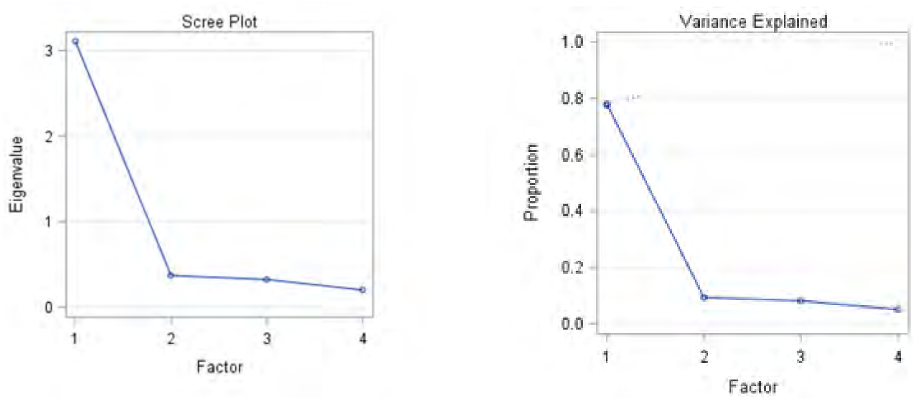
For the overall sample, each student's 34 Star Math item responses were divided into subsets of items aligned to each of the 4 blueprint domains. Tests administered in grades K–8 included items from all four domains. Tests given in grades 9–12 included items from just 3 domains with no items measuring data analysis, probability and statistics domain.

For each student, separate Rasch ability estimates (subtest scores) were calculated from each domain-specific subset of item responses. A Bayesian sequential procedure developed by Owen (1969, 1975) was used for the subtest scoring. Across all grade bands, the number of items included in each math subtest ranged from 3 to 23 items for the NUM domain, 1 to 18 items for the ALG domain, 5 to 13 items for the GEO domain, and 0 to 3 items for the DAT domain, following the Star Math test blueprints, which specify different numbers of items per domain, depending on the student's grade level.

Intercorrelations of the blueprint domain-specific Rasch subtest scores were analyzed using exploratory factor analysis (EFA) to evaluate the number of dimensions/ factors underlying Star Math domain scores. In each grade band, the

EFA analyses retained a single dominant underlying dimension based on either the MINEIGEN (eigenvalue greater than 1) or the PROPORTION criterion (proportion of variance explained by the factor), as expected. An example of a scree plot from grade band K to 2 based on the PROPORTION criterion is shown in Figure 3. Similar scree plots showing a single dominant factor for the first eigenvalue and extracted factor were found at all grade bands and across grade bands. EFA analyses using both SAS and SPSS software showed one significant factor at each grade band and across all grade bands for principal components analysis, unweighted least squares factors, generalized least squares factors, maximum likelihood factors, alpha factors, image factors. Standardized factor loadings for each domain were always above 0.80 for the first extracted factor.

Figure 3: Example Scree and Variance Explained Plots from the Grade Band K to 2 Exploratory Factor Analysis in Star Math



Confirmatory factor analyses (CFA) were also conducted using the subtest scores from the CFA analysis. A separate confirmatory analysis was conducted for each grade band. The CFA models tested a single underlying model as shown in Figure 4. One CFA model with four domains was fitted for students in grade bands K to 2, 3 to 5, and 6 to 8; a second CFA model with three domains was fitted for students in grade band 9 to 12 since the test blueprint did not administer items from the domain for Data Analysis, Probability and Statistics.

Figure 4: Confirmatory Factor Analyses (CFA) in Star Math

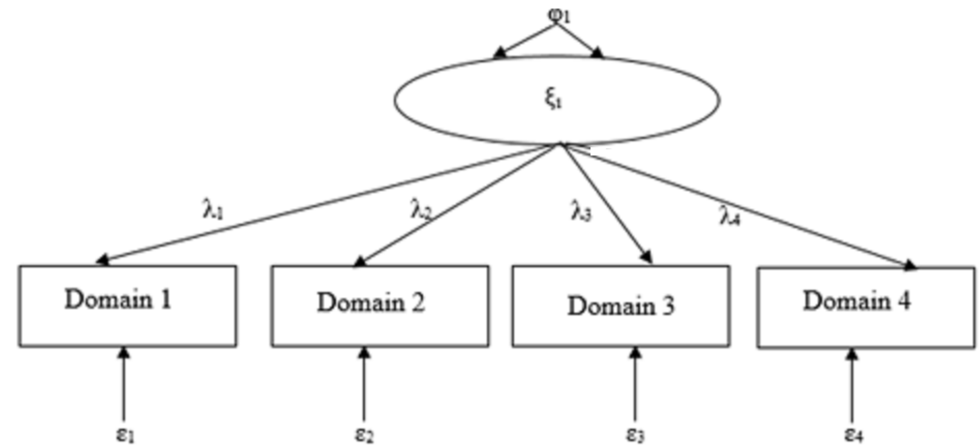


Table 20: Domain Scores Included in the CFA Models for Star Math, by Grade Band[a]

Grade Bands	Domains			
	1	2	3	4
K to 2	ALG	GEO	DAT	NUM
3 to 5	ALG	GEO	DAT	NUM
6 to 8	ALG	GEO	DAT	NUM
9 to 12	ALG	GEO		NUM

a. Math Domain Key:
ALG = Algebra Domain
GEO = Geometry and Measurement Domain
DAT = Data Analysis, Statistics, and Probability Domain
NUM = Numeration and Operations Domain

The results of the CFA analyses by grade band and across all grade bands are summarized in Table 21 on page 58. Grade Band ALL4 shows results across all grade bands for four math domains (ALG, GEO, DAT, and NUM); Grade Band ALL3 shows results across all grade bands for three math domains (ALG GEO and NUM). The CFA models for Grade band 9 to 12 and for Grade band ALL3 were just-identified statistical models and required fixing the expected error variance for one estimated analysis parameter. The analyst fixed the error variance for the NUM domain at its computed value for these analyses, since the NUM domain had the least number of blueprint specified items for grade band 9 to 12 which also affected estimation of grade band ALL3 at the high school level.

Table 21: Summary of the Goodness-of-Fit of the CFA Models for Star Math by Grade Band

Grade Band	N	χ^2	df	CFI	GFI	NFI	RMSEA	SRMR
K to 2	35,216	95.8252	2	0.9990	0.9986	0.9990	0.0365	0.0055
3 to 5	32,095	26.0192	2	0.9997	0.9996	0.9997	0.0193	0.0025
6 to 8	47,477	165.487	2	0.9989	0.9983	0.9989	0.0415	0.0047
9 to 12	69,133	354.493	1	0.9977	0.9966	0.9977	0.0715	0.0139
ALL4	131,221	173.930	2	0.9997	0.9993	0.9997	0.0256	0.0014
ALL3	201,088	584.864	1	0.9991	0.9980	0.9869	0.0539	0.0032

As Table 21 indicates, sample sizes ranged from 32,095 to 69,133 within grade bands; because the chi-square (χ^2) test is not a reliable test of model fit when sample sizes are large, a variety of fit indices are presented. The comparative fit index (CFI), goodness of fit index (GFI), and the normed fit index (NFI) are shown; for these indices, values are either 1 or very close to 1, indicating strong evidence of a single construct/dimension for Star Math. In addition, the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) are presented. RMSEA and SRMR values less than 0.08 indicate good fit. Cutoffs for the indices are presented in Hu and Bentler, 1999. Overall, the CFA results strongly support a single underlying dimension.

Table 22 presents the CFA Factor Loadings for the four math content domains for Algebra (ALG), Geometry and Measurement (GEO), Data Analysis, Statistics and Probability (DAT) and Numeration and Operations (NUM). These results show consistently high factor loadings within grade bands across the three to four math domains, and across grade bands within each math domain cluster. The CFA factor loading range from 0.78 to 0.93 show congruence of factor loadings within domains across grade bands, and within grade bands across the math content domains. Grade Band ALL4 shows results across all grade bands for four math domains (ALG, GEO, DAT, and NUM). Grade Band ALL3 shows results across all grade bands for three math domains (ALG, GEO, and NUM).

Table 22: Summary of the CFA Factor Loadings for Star Math by Grade Band and Math Domain

	CFA Factor Loadings			
Grade Band	ALG	GEO	DAT	NUM
K to 2	0.7799	0.8581	0.7994	0.9168
3 to 5	0.8104	0.8682	0.8213	0.9321
6 to 8	0.8681	0.8652	0.8259	0.9350
9 to 12	0.9275	0.9254	NA*	0.8099
ALL4	0.9320	0.9400	0.9165	0.9690
ALL3	0.9465	0.9523	NA*	0.9356

Table 23 summarizes principal components and principal axis Exploratory Factor analysis (EFA) factor loadings for Star Math domains across all grade bands. These results show independent support of the CFA analyses' results for the unidimensionality for Star Math. Note, the component and factor loadings for the DAT math domain are estimated from grades K to 8 but were not available for grades 9 to 12 due to the test blueprint.

Table 23: Summary of Principal Components and Principal Axis EFA Factor Loadings Across All Grade Bands for Star Math Domains

Principal Components		Principal Axis
	Component	Factor
Math Domain	1	1
ALG	0.951	0.932
DAT	0.944	0.918
GEO	0.955	0.940
NUM	0.970	0.968

The EFA analyses were conducted using the factor procedure in SAS 9.4 software and in IBM SPSS version 19 software, while the CFA analysis was conducted using the calis procedure in the SAS 9.4 software (SAS Institute, Cary NC).

Relationship of Star Math Scores to Scores on Other Tests of Mathematics Achievement

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

During a subsequent norming of Star Math, scores on other standardized tests were obtained for more than 30,000 additional students. All of the standardized tests listed above were included, plus others such as Northwest Evaluation Association (NWEA) and TerraNova. Scores on state assessments from the following states were also included: Arkansas, Connecticut, Delaware, Florida, Georgia, Kentucky, Idaho, Indiana, Illinois, Maryland, Michigan, Minnesota, Mississippi, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Dakota, Texas, Virginia, and Washington. The extent that the Star Math test correlates with these tests provides support for its construct validity. That is, strong and positive correlations between Star Math and these other instruments provide support for the claim that Star Math effectively measures mathematics achievement.

Table 24, Table 25, Table 26, and Table 27 (starting on page 63) summarize the correlation coefficients between the scores on the Star Math test and each of the other test instruments for which data were received. "Appendix B: Additional Evidence of Star Math Validity" on page 150 contains detailed correlational data behind the summaries in these tables.

Table 24 and Table 25 summarize "concurrent validity" data, that is, correlations between Star Math norming study test scores and other tests administered within a two-month time period.

In addition to the concurrent validity estimates summarized in Table 24 and Table 25, data concerning Star Math's predictive validity are summarized in Table 26 and Table 27. Predictive validity provides an estimate of the extent to which scores on the Star Math test predicted scores on criterion measures given at a later point in time, operationally defined as more than 2 months between the Star test (predictor) and the criterion test. It provides an estimate of the linear relationship between Star scores and scores on measures covering a similar academic domain. Predictive correlations are typically attenuated by time due to the fact that students are gaining skills in the interim between testing occasions, and also by differences between the tests' content specifications.

The following is a partial list of math assessments for which there is evidence of correlations with Star Math reported in this technical manual.

- Achievement level (RIT) Test
- ACT Aspire
- American College Testing Program
- Arkansas Augmented Benchmark Examination (AABE)
- California Achievement Test
- Canadian Achievement Test
- Cognitive Abilities Test
- Comprehensive Test of Basic Skills
- Connecticut Mastery Test
- Delaware Student Testing Program (DSTP)
- Des Moines Public School (Grade 2 pretest)
- Differential Aptitude Tests

- Educational Development Series
- Explore Tests
- Florida Comprehensive Assessment Test (FCAT)
- Florida Standards Assessments (FSA)
- Georgia Milestones
- Georgia High School Graduation Test
- Idaho Standards Achievement Test (ISAT)
- Indiana Statewide Testing for Educational Progress
- Iowa Assessment
- Iowa Test of Basic Skills (ITBS)
- Kansas State Assessment Program (KSAP)
- Kentucky Core Content Test (KCCT)
- Kentucky Core Content Test
- Key Stage 2 Standardised Attainment Tests (UK KS2 SATs)
- Maryland High School Placement Test
- McGraw Hill Mississippi/Criterion Referenced
- Metropolitan Achievement Test (MAT)
- Michigan Educational Assessment Program
- Minnesota Comprehensive Assessment (MCA)
- Mississippi Academic Assessment Program (MAAP)
- Mississippi Curriculum Test (MCT2)
- Missouri Assessment Program (MAP) Grade-Level Tests
- Multiple Assessment Series (Primary Grades)
- New Jersey Assessment of Skills and Knowledge (NJASK)
- New Standards Reference Mathematics Exam (Rhode Island)
- New York State Assessment Program
- New York State Math Assessment
- North Carolina End-of-Grade (NCEOG) Test
- Northwest Evaluation Association Levels Test
- NWEA, NALT, & MAP
- Ohio Achievement Assessment
- Ohio Proficiency Test (OPT)
- Ohio State Tests (OST)

- Oklahoma Core Curriculum Test (OCCT)
- Oklahoma School Testing Program Core Curriculum Tests
- Oregon State Assessment
- Otis Lennon School Ability Test (OLSAT)
- Palmetto Achievement Challenge Test (PACT), 2001
- Partnership for Assessment of Readiness for College and Careers (PARCC)
- Pennsylvania System of School Assessment (PSSA)
- PLAN
- Preliminary SAT/National Merit Scholarship Qualifying Test
- Smarter Balanced Assessment (SBA)
- South Dakota State Test of Educational Progress (DSTEP)
- Stanford Achievement Test
- Star Math
- State of Texas Assessments of Academic Readiness Standards Test 2
- Tennessee Comprehensive Assessment Program (TCAP)
- TerraNova
- Test of Achievement Proficiency
- Test of New York State Standards
- Texas Assessments of Academic Readiness Standards
- Texas Assessment of Academic Skills (TAAS), 2001
- Texas Assessment of Knowledge and Skills (TAKS)
- Transitional Colorado Assessment Program (TCAP)
- Virginia Standards of Learning
- Washington Assessment of Student Learning
- West Virginia Educational Standards Test 2
- Wide Range Achievement Test
- Wisconsin Forward Exam
- Wisconsin Knowledge and Concepts Examination (WKCE)

Table 24, Table 25, Table 26, and Table 27 contain summaries of some of the correlational data in support of Star Math validity. Table 24 summarizes the within-grade average concurrent validity coefficients for grades 1–6; these varied from 0.64–0.75, with an overall average of 0.73. Table 25 summarizes the concurrent validity for grades 7–12; correlations ranged from 0.56–0.74, with an overall average of 0.71.

Table 26 and Table 27 contain similar summaries of predictive validity coefficients. Table 26 summarizes the grades 1–6 data; coefficients ranged from 0.55–0.74, with an average of 0.72. Table 27 does the same for grade 7–12 predictive validity; obtained coefficients ranged from 0.65–0.77, with an average of 0.74.

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

Table 24: Summary of Concurrent Validity Statistics for Grades 1–6: Star Math Correlations (r) with External Tests Administered Between 2002 and 2016

Summary							
Grade(s)	**Total N**	**1**	**2**	**3**	**4**	**5**	**6**
Number of students	370,651	215	951	104,603	99,768	93,810	71,304
Number of coefficients	241	5	11	64	56	62	43
Average validity	–	0.65	0.64	0.72	0.73	0.75	0.72
Overall average	0.73						

Table 25: Summary of Concurrent Validity Statistics for Grades 7–12: Star Math Correlations (r) with External Tests Administered Between 2002 and 2016

Summary							
Grade(s)	**Total N**	**7**	**8**	**9**	**10**	**11**	**12**
Number of students	123,819	60,917	51,442	5,335	4,528	1,494	103
Number of coefficients	95	36	36	5	7	6	5
Average validity	–	0.73	0.74	0.65	0.58	0.70	0.56
Overall average	0.71						

Table 26: Summary of Predictive Validity Data, Grades 1–6: Star Fall-to-Spring Correlations (r) with External Tests Administered Between 2001 and 2016

Summary							
Grade(s)	**Total N**	**1**	**2**	**3**	**4**	**5**	**6**
Number of students	662,040	11,880	33,076	176,784	175,330	152,693	112,277
Number of coefficients	285	6	10	77	69	74	49
Average validity	–	0.55	0.63	0.72	0.74	0.73	0.74
Overall average	0.72						

Table 27: Summary of Predictive Validity Data, Grades 7–12: Star Fall-to-Spring Correlations (r) with External Tests Administered Between 2001 and 2016

Summary							
Grade(s)	**Total N**	**7**	**8**	**9**	**10**	**11**	**12**
Number of students	160,323	75,876	59,960	7,971	8,708	6,831	977
Number of coefficients	126	51	46	8	9	9	3
Average validity	–	0.75	0.74	0.75	0.77	0.72	0.65
Overall average	0.74						

Meta-Analysis of the Star Math Validity Data

Meta-analysis is a set of statistical procedures that combines results from different sources or studies. When applied to a set of correlation coefficients that estimate test validity, meta-analysis combines the observed correlations and sample sizes to yield estimates of overall validity, as well as standard errors and confidence intervals, both overall and within grades.

To conduct a meta-analysis of the Star Math validity data, the 747 correlations summarized in Table 24, Table 25, Table 26, and Table 27, observed in data from Star Math tests of more than 1.3 million students, were combined and analyzed using a fixed effects model for meta-analysis. The results are displayed in Table 28. The table lists results for the correlations within each grade, as well as results with all twelve grades' data combined. For each set of results, the table lists an estimate of the true validity, a standard error, and the lower and upper limits of a 95 percent confidence interval for the validity coefficient. Based on the 747 correlation coefficients, the overall estimate of the validity of Star Math is 0.758, with a standard error of 0.001. The probability of observing the 747 correlations reported in Table 24, Table 25, Table 26, and Table 27, if the true validity were zero, is virtually zero. Because the correlations were obtained with widely different tests, and among students from twelve different grades, these results provide strong support for the validity of Star Math as a measure of math skills.

Table 28: Results of the Meta-Analysis of Star Math Correlations with Other Tests

Grade	Effect Size		95% Confidence Interval		Total Correlations	Total N
	Validity Estimate	Standard Error	Lower Limit	Upper Limit		
1	0.558	0.009	0.545	0.570	11	12,095
2	0.627	0.005	0.620	0.633	21	34,027
3	0.755	0.002	0.753	0.756	141	281,387
4	0.760	0.002	0.759	0.762	125	275,098
5	0.765	0.002	0.764	0.767	136	246,503
6	0.777	0.002	0.775	0.779	92	183,581
7	0.770	0.003	0.768	0.772	87	136,793
8	0.754	0.003	0.751	0.756	82	111,402
9	0.708	0.009	0.699	0.716	13	13,306
10	0.751	0.009	0.744	0.759	16	13,236
11	0.740	0.011	0.730	0.750	15	8,325
12	0.731	0.030	0.702	0.758	8	1,080
All Grades	0.758	0.001	0.757	0.759	747	1,316,833

Linking Star and State Assessments: Comparing Student- and School-Level Data

With an increasingly large emphasis on end-of-the-year summative state tests, many educators seek out informative and efficient means of gauging student performance on state standards—especially those hoping to make instructional decisions before the year-end assessment date.

For many teachers, this is an informal process in which classroom assessments are used to monitor student performance on state standards. While this may be helpful, such assessments may be technically inadequate when compared to more standardized measures of student performance. Recently the assessment scale associated with Star Math has been linked to the scales used for summative mathematics tests in most states. Linking Star Math assessments to state tests allows educators to reliably predict student performance on their state assessment using Star Math scores. More specifically, it places teachers in a position to identify

- which students are on track to succeed on the year-end summative state test, and
- which students might need additional assistance to reach proficiency.

Educators using Star Math assessments can access Star Performance Reports that allow access to students' Pathway to Proficiency. These reports indicate whether individual students or groups of students (by class, grade, or demographic characteristics) are likely to be on track to meet a particular state's criteria for mathematics proficiency. In other words, these reports allow instructors to evaluate student progress toward proficiency and make data-based instructional decisions well in advance of the annual state tests. Additional reports automatically generated by Star Math help educators screen for later difficulties and progress monitor students' responsiveness to interventions.

Relationship of Star Math Scores to Scores on Multi-State Consortium Tests in Math

In recent years, the National Governors' Association, in collaboration with the Council of Chief State School Officers (CCSSO), developed a proposed set of curriculum standards in English Language Arts and Math, called the Common Core State Standards. Forty-five states voluntarily adopted those standards; subsequently, many states have dropped them, but several states continue to use them or base their own state standards on them. Two major consortia were formed to develop assessments systems that embodied those standards: the Smarter Balanced Assessment Consortium (SBAC) and Partnership for Assessment of Readiness for College and Careers (PARCC). SBAC and PARCC end-of-year assessments have been administered in numerous states in place of those states' previous annual accountability assessments. Renaissance Learning was able to obtain SBAC and PARCC scores of many students who had taken Star Math earlier in the same school years. Table 29 and Table 30 contain coefficients of correlation between Star Math and the consortium tests. The average of the concurrent correlations was approximately 0.88 for SBAC and 0.83 for PARCC. The average predictive correlation was approximately 0.89 with the SBAC assessments, and 0.85 for PARCC.

Table 29: Concurrent and Predictive Validity Data: Star Math Scaled Scores Predicting Later Performance for Grades 3–8 on Smarter Balanced Assessment Consortium Tests

Star Math Concurrent and Predictive Correlations with Smarter Balanced Assessment Scores							
Grade		**3**	**4**	**5**	**6**	**7**	**8**
Concurrent	N	10,800	10,582	9,750	7,852	6,344	5,424
	Correlation	0.86	0.88	0.89	0.87	0.88	0.87
Predictive	N	8,593	8,571	8,595	8,575	8,623	8,859
	Correlation	0.89	0.90	0.90	0.89	0.89	0.86

Table 30: Concurrent and Predictive Validity Data: Star Math Scaled Scores Correlations for Grades 3–8 with PARCC Assessment Consortium Test Scores

Star Math Concurrent and Predictive Correlations with PARCC Assessment Scores							
Grade		**3**	**4**	**5**	**6**	**7**	**8**
Concurrent	N	3,635	4,008	3,653	4,150	4,066	3,748
	Correlation	0.83	0.86	0.82	0.83	0.81	0.80
Predictive	N	4,103	4,787	4,266	5,050	4,368	4,196
	Correlation	0.83	0.82	0.78	0.79	0.80	0.77

Classification Accuracy of Star Math

Accuracy for Predicting Proficiency on a State Math Assessment

Star Math test scores have been linked statistically to numerous state Math assessment scores. The linked values have been employed to use Star Math to predict student proficiency in Math on those state tests. One example of this is a linking study conducted using a multi-state sample of students' scores on the PARCC consortium assessment.[1] Table 31 presents classification accuracy statistics for grades 3 through 8.

Table 31: Classification Diagnostics for Predicting Students' Math Proficiency on the PARCC Consortium Assessment from Earlier Star Math Scores

Measure	Grade					
	3	**4**	**5**	**6**	**7**	**8**
Overall classification accuracy	89%	90%	92%	91%	91%	90%
Sensitivity	71%	58%	57%	66%	59%	59%
Specificity	94%	97%	98%	97%	97%	96%
Positive predictive value (PPV)	75%	82%	83%	79%	77%	77%
Negative predictive value (NPV)	93%	91%	93%	93%	93%	92%
Observed proficiency rate (OPR)	20%	19%	15%	17%	16%	17%
Projected proficiency rate (PPR)	19%	13%	10%	14%	12%	13%
Proficiency status projection error	–1%	–6%	–5%	–3%	–4%	–4%
Area Under the ROC Curve	0.94	0.94	0.94	0.95	0.95	0.94

1. Renaissance Learning (2016). Relating Star Reading® and Star Math® to the Colorado Measure of Academic Success (CMAS) (PARCC Assessments) Performance.

As the table shows, overall classification accuracy ranged from 89% to 92%, depending on grade. Area Under the Curve (AUC) was at least 0.94 for all grades. Specificity was especially high, and the projected proficiency rates were very close to the observed proficiency rates at all grades.

Numerous other reports of linkages between Star Math and state accountability tests have been conducted. Reports are available at research.renaissance.com/.

Evidence of Technical Adequacy for Informing Screening and Progress Monitoring Decisions

Many school districts use tiered models such as Response to Intervention (RTI) or Multi-Tiered Systems of Support (MTSS) to guide instructional decision making and improve outcomes for students. These models represent a more proactive, data-driven approach for better serving students as compared with prior decision-making practices, including processes to:

- Screen all students to understand where each is in the progression of learning in reading, math, or other disciplines
- Identify at-risk students for intervention at the earliest possible moment
- Intervene early for students who are struggling or otherwise at-risk of falling behind; and
- Monitor student progress in order to make decisions as to whether they are responding adequately to the instruction/intervention

Assessment data are central to both screening and progress monitoring, and Star Math is widely used for both purposes. This chapter includes technical information about Star Math's ability to accurately screen students according to risk and to help educators make progress monitoring decisions. Much of this information has been submitted to and reviewed by the Center on Response to Intervention https://rti4success.org/ and/or the National Center on Intensive Intervention https://intensiveintervention.org/, two technical assistance groups funded by the US Department of Education.

For several years running, Star Math has enjoyed favorable technical reviews for its use in informing screening and progress monitoring decision by the CRTI and NCII, respectively. The most recent reviews by CRTI indicate that Star Math has a "convincing" level of evidence (the highest rating awarded) in the core screening categories, including classification accuracy, reliability, and validity. CRTI also notes that the extent of the technical evidence is "Broad" (again, the highest rating awarded) and notes that not only is the overall evidence compelling, but there are disaggregated data as well that shows Star Math works equally well among subgroups. The most recent reviews by NCII indicate that there is fully “convincing” evidence of Star Math’s psychometric quality for progress monitoring purposes, including reliability, validity, reliability of the slope, and validity of the slope. Furthermore, they find fully “convincing” evidence that Star Math is sufficiently sensitive to student growth, has adequate alternate forms, and provides data-based guidance to educators on end-of-year benchmarks and when an intervention should be changed, among other categories. Readers may find additional information on Star Math on those sites and should note that the reviews are updated on a regular basis, as their review standards are adjusted and new technical evidence for Star Math and other assessments are evaluated.

Screening

According to the Center on Response to Intervention, “Screening is conducted to identify or predict students who may be at risk for poor learning outcomes. Universal screening assessments are typically brief, conducted with all students at a grade level, and followed by additional testing or short-term progress monitoring to corroborate students’ risk status.”[2]

Most commonly, screening is conducted with all students at the beginning of the year and then another two to four times throughout the school year. Star Math is widely used for this purpose. In this section, the technical evidence supporting its use to inform screening decisions is summarized.

Organizations of RTI/MTSS experts such as the Center on Response to Intervention and the RTI Action Network[3] are generally consistent in how measurement tools should be evaluated for their appropriateness as screeners. Key categories include the following:

1. **Validity and reliability**. See the “Reliability and Measurement Precision” chapter and the earlier sections of this “Validity” chapter for a summary of the available evidence supporting Star Math’s reliability and validity.
2. **Practicality and efficiency.** Screening measures should not require much teacher or student time. Because most students can complete a Star Math test in 15–20 minutes or less, and because it is group administered and scored automatically, Star Math is an exceptionally efficient general outcomes measure for mathematics.
3. **Classification accuracy metrics** including sensitivity, specificity, and overall predictive accuracy. These are arguably the most important indicators, addressing the main purpose of screening: When a brief screening tool

2. https://rti4success.org/essential-components-rti/universal-screening
3. http://www.rtinetwork.org/learn/research/universal-screening-within-a-rti-model

indicates a student either is or is not at risk of later difficulties in mathematics, how often is it accurate, and what types of errors are made?

It is common to use high-stakes indicators such as state summative assessments as criterion measures for classification accuracy evaluation. Star Math is linked to virtually every state summative assessment in the US as well as the United Kingdom's Key Stage 2 Standardised Attainment Tests for Maths, as well as the ACT and SAT college entrance exams. The statistical linking of the Star Math scale with these other measures' scales, combined with Star Math growth norms (discussed in the Growth Norms section, on page 81 of the Norming chapter) empowers Star Math reports, dashboards, and data extracts to make predictions throughout the school year about future student performance. These predictions inform educator screening decisions in schools using an RTI/MTSS framework. (Educators are also free to use norm-referenced scores such as Percentile Ranks to inform screening decisions.)

Star Math's classification accuracy results from several recent predictive studies are summarized in Table 32 on page 71. Each study evaluated the extent to which Star Math accurately predicted whether a student achieved a specific performance level on another mathematics measure. The specific performance level (cut point) varies by assessment and grade. Cut points are set by assessment developers and sponsors, which in the case of state summative exams usually means the state department of education and/or state board of education. State assessments generally have between three and five performance levels, and the cut point used in these analyses refers to the level the state has determined indicates meeting grade level mathematics standards. For instance, the cut point on California's CAASPP is Level 3, also known as "Standard Met." On Louisiana's LEAP 2025 the cut point is at the "Mastery" level. In the case of ACT and SAT, the cut point established by the developers (ACT and College Board, respectively) indicates an estimated level of readiness for success in college.

Table 32: Summary of Classification Accuracy Metrics from Recent Studies Linking Star Math with Summative Mathematics Measures

Assessment	Grade/s Covered	Date Study Completed	Study Sample Size	Average Result Across All Grades: Overall Classification Accuracy	Sensitivity	Specificity	Area under ROC Curve
ACT Mathematics (college readiness)	11	4/22/2016	6,328	89%	67%	98%	0.93
ACT Aspire	3–8	6/1/2017	37,581	85%	81%	80%	0.92
California Assessment of Student Performance and Progress (CAASPP) (Smarter Balanced)	3–8	10/30/2015	51,816	87%	84%	88%	0.94
Florida Standards Assessments (FSA)	3–8	6/30/2015	16,071	83%	83%	81%	0.91
Georgia Milestones	3–8	7/1/2017	44,745	89%	77%	93%	0.94
Illinois Partnership for Assessment of Readiness for College and Careers (PARCC) Assessments	3–8	7/13/2016	23,260	91%	62%	96%	0.94
Louisiana Educational Assessment Program (LEAP 2025)	3–8	1/31/2018	7,713	84%	73%	87%	0.91
Maine Educational Assessment (MEA)	3–8	7/1/2017	895	86%	78%	88%	0.91
Mississippi Academic Assessment Program (MAAP)	3–8	2/1/2017	10,954	85%	78%	88%	0.92
Missouri Assessment Program (MAP) Grade-Level Tests	3–8	3/14/2017	19,442	84%	79%	86%	0.94
North Carolina READY End-of-Grade (EOG)	3–8	2/16/2015	125,932	81%	78%	82%	0.89
Ohio State Tests	3–8	12/20/2016	19,682	83%	79%	86%	0.92
Pennsylvania’s System of School Assessment (PSSA)	3–8	12/19/2016	3,436	87%	87%	86%	0.94
SAT (college entrance)	11	10/3/2018	2,126	84%	70%	87%	0.91
South Carolina College- and Career-Ready Assessments (SC READY)	3–8	12/5/2016	8,909	87%	83%	89%	0.94

Table 32: Summary of Classification Accuracy Metrics from Recent Studies Linking Star Math with Summative Mathematics Measures

Assessment	Grade/s Covered	Date Study Completed	Study Sample Size	Average Result Across All Grades			
				Overall Classification Accuracy	Sensitivity	Specificity	Area under ROC Curve
State of Texas Assessments of Academic Readiness (STAAR)	3–7	7/1/2017	642	84%	80%	85%	0.91
State of Texas Assessments of Academic Readiness (STAAR) Algebra 1 End of Course (EOC) Test	Algebra I	2/9/2017	3,292	76%	85%	60%	0.82
UK Key Stage 2 Standardised	Year 6	9/1/2017	815	89%	89%	90%	0.97
Attainment Tests (SATs) Maths							
Wisconsin Forward Exam	3–8	12/22/2016	39,812	91%	71%	96%	0.96

Notes:

- Some tests, such as the Smarter Balanced (indicated above for California) and PARCC (indicated above for Illinois) are used in multiple states, so those results may apply to other states not listed here.
- Overall classification accuracy refers to the percentage of correct classifications.
- Sensitivity refers to the rate at which Star Math identifies students as being at-risk who demonstrate a poor learning outcome at a later point in time. Sensitivity can be thought of as the true positive rate. Screening tools with high sensitivity help ensure that students who truly need intervention will be identified to receive it.
- Specificity refers to the rate at which Star Math identifies students as being not at-risk who perform satisfactorily at a later point in time. Specificity can be thought of as a true negative rate. Screening tools with high specificity help ensure that scarce resources are not invested in students who do not require extra assistance.
- Area under the ROC (Receiver Operating Characteristic) curve is a powerful indicator of overall accuracy. The ROC curve is a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity) for the full range of possible screener (Star Math) cut points. The area under ROC Curve (AUC) is

an overall indication of the diagnostic accuracy of the curve. AUC values range between 0 and 1 with 0.5 indicating a chance level of accuracy. The Center for Response to Intervention considers results at or above 0.85 to be an indication of convincing evidence of classification accuracy.[4]

Note that many states tend to not use the same assessment system for more than a few consecutive years, and Renaissance endeavors to keep the Star Math classification reporting as up to date as possible. Those interested in reviewing the full technical reports for these or other state assessments are encouraged to visit http://research.renaissance.com/advancedsearch.asp and search by state name for the Star Math linking reports (e.g., “Wisconsin linking”).

Progress Monitoring

According to the National Center on Intensive Intervention, “progress monitoring is used to assess a student’s performance, to quantify his or her rate of improvement or responsiveness to intervention, to adjust the student’s instructional program to make it more effective and suited to the student’s needs, and to evaluate the effectiveness of the intervention.”[5]

In an RTI/MTSS context, progress monitoring involves frequent assessment—usually occurring once every 1–4 weeks—and often involves only those students who are receiving additional instruction after been identified as at-risk via the screening process. Ultimately, educators use progress monitoring data to determine whether a student is responding adequately to the instruction, or whether adjustments need to be made to the instructional intensity or methods. The idea is to get to a decision quickly, with as little testing as possible, so that valuable time is not wasted on ineffective approaches. Educators make these decisions by comparing their performance against a goal set by the educator. Goals should be “reasonable yet ambitious”[6] as recommended by Shapiro (2008), and Star Math offers educators a variety of guidance to set normative or criterion-referenced goals that meet these criteria.

The RTI Action Network, National Center on Intensive Intervention, and other organizations offering technical assistance to schools implementing RTI/MTSS models are generally consistent in encouraging educators to select assessments for progress monitoring that have certain characteristics. A summary of those characteristics and relevant information about Star Math is provided below.

4. https://rti4success.org/resources/tools-charts/screening-tools-chart/screening-tools-chart-rating-system
5. https://intensiveintervention.org/ncii-glossary-terms#Progress Monitoring
6. Shapiro, E. S. (2008). Best practices in setting progress-monitoring monitoring goals for academic skill improvement. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology V (pp. 141–157). Bethesda, MD: National Association of School Psychologists.

1. **Evidence of psychometric quality.**

   a. **Reliability and validity.** See the "Reliability and Measurement Precision" chapter and the earlier sections of this "Validity" chapter for a summary of the available evidence supporting Star Math's reliability and validity.

   b. **Reliability of the slope.** Because progress monitoring decisions often involve the student's rate of progress over multiple test administrations, the characteristics of the student's slope of improvement, or trend line, are also important. A study was conducted in 2017 by Renaissance Learning to evaluate reliability of slope for at-risk students who were being progress monitored during the 2016–17 school year. Specifically, the sample included 96,209 students who began the school year at-risk (defined as placing below the 30th Percentile Rank in Star Math) and were assessed 10 or more times during the school year, with a minimum of 140 days between first and last test.

      Every student's Star Math test records were sorted in chronological order. Each test record was coded as either an odd- or even-numbered test. Slopes were estimated for each student's odd-number tests and also for the even-numbered tests using ordinary least squares regression. Then, the odd and even slopes were correlated. The table below summarizes the Pearson correlation coefficients by grade, indicating a consistently strong association between even and odd numbered test slopes.

Table 33: Star Math Reliability of the Slope Coefficients by Grade, 1–12

Grade	n	Coefficient
1	8,987	0.92
2	18,460	0.93
3	16,696	0.93
4	14,738	0.93
5	12,411	0.93
6	8,627	0.94
7	6,379	0.93
8	5,317	0.93
9	2,129	0.94
10	1,265	0.94
11	803	0.94
12	397	0.94

2. **Produce a sufficient number of forms.** Because Star Math is computer-adaptive with an item bank comprising more than six thousand items, there are at a minimum, several hundred alternate forms for a student at a given ability level. This should be more than sufficient for even the most aggressive progress monitoring testing schedule.

   A variety of grade-specific evidence is available to demonstrate the extent to which Star Math can reliably produce consistent scores across repeated administrations of the same or similar tests to the same individual or group. These include:

   a. Generic reliability, defined as the proportion of test score variance that is attributable to true variation in and the trait or construct the test measures. Grade-level results are summarized in Table 12 on page 45 and Table 16 on page 53.

   b. Alternate form reliability, defined as the correlation between test scores on repeated administrations to the same examinees. Grade-level results are summarized in Table 12 on page 45 and Table 16 on page 53.

   c. **Practicality and efficiency.** As mentioned above, most students complete Star Math in 15–20 minutes. It is auto-scored and can be group administered, requiring very little educator involvement, making it an efficient progress monitoring solution.

3. **Specify criteria for adequate growth and benchmarks for end-of-year performance levels.** Goal-setting decisions are handled by local educators, who know their students best and are familiar with the efficacy and intensity of the instructional supports that will be offered. That said, publishers of assessments used for progress monitoring are expected to provide empirically based guidance to educators on setting goals.

   Star Math provides guidance to inform goal setting using a number of different metrics, including the following:

   a. **Student Growth Percentile.** SGP describes a student's velocity (slope) relative to a national sample of academic peers—those students in the same grade with a similar score history. SGPs work like Percentile Ranks (1–99 scale) but once an SGP goal has been set, it is converted to a Scaled Score goal at the end date specified by the teacher. An SGP-defined goal can be converted into an average weekly increase in a Scaled Score metric if educators prefer to use that. Many teachers select either SGP 50 (indicating typical or expected growth) as minimum acceptable growth, or something indicating accelerated growth, such as 65 or 75. A helpful feature of

SGP is that it can be used as a “reality check” for any goal, whether it be in an SGP metric or something else (e.g., Scaled Score, Percentile Rank). SGP estimates the likelihood that the student will achieve a level of growth or later performance. For example, a goal associated with an SGP of 75 indicates that only about 25 percent of the student’s academic peers would be expected to achieve that level of growth.

b. **State test proficiency.** As described in the Screening section, the fact that Star Math is linked to virtually every state assessment enables educators to select values on the Star scale that are approximately equivalent to states’ defined proficiency level cut points for each grade.

c. **Percentile Rank and Scaled Score.** Educators may also enter custom goals using Percentile Rank or Scaled Score metrics.

Additional Research on Star Math as a Progress Monitoring Tool

A 2016 study by Cormier & Bulut[7] evaluated Star Math as a progress monitoring tool, concluding:

- Although relatively little research exists on using computer adaptive measures for progress monitoring as opposed to curriculum based measurement probes, the study concluded it was possible to use Star Math for progress monitoring purposes.
- Sufficiently reliable progress monitoring slopes could be generated in as few as five Star Math administrations.
- The duration of Star Math progress monitoring (i.e., over how many weeks should be conducted) is a function of the amount of typical growth by grade in relation to measurement error. For earlier grades (when student rates of growth are greatest), that amount of time could be as little as six weeks. For middle grades, 20 weeks should be sufficient.
- These two findings challenge popular rules of thumb about progress monitoring frequency and duration (most of which are derived from CBM probe studies), which often involve weekly testing over periods of time that are selected due to popular convention rather than empirical evidence.
- Using Theil-Sen regression procedures to estimate slope as opposed to OLS could reduce the influence of outlier scores, and thus provide a more accurate picture of student growth.

7. Cormier, D. & Bulut, O. (2016). *Developing psychometrically sound decision rules for Star Math.* Report prepared for Renaissance Learning.

Differential Item Functioning

Ensuring that an assessment is not biased against different demographic subgroups that take the assessment is a fundamental aspect of showing test fairness and providing validity evidence to support the interpretations and uses of the assessment. One strategy that is often used as part of evaluating test fairness is a strategy known as differential item functioning (DIF). DIF occurs when two or more demographic subgroups perform differently on an item after controlling for performance on the test (Holland & Thayer, 1988; Zumbo, 2007). In other words, for students of similar ability, an item that displays DIF may appear to favor one group of students based on demographics such as gender and/or race/ethnicity.

There are many different methods that one can use to investigate items for DIF, including item response theory methods, observed score methods, and a variety of nonparametric approaches (Zumbo, 2007). Star Math uses the logistic regression (LR) method to evaluate items for DIF (Rogers & Swaminathan, 1993; Swaminathan & Rogers, 1990; Swaminathan, 1994). With this approach, student item responses are regressed on student ability estimates from Star Math as well as their subgroup membership and the student ability and subgroup membership interaction. To conduct a DIF analysis, a reference group and a focal group is defined. For instance, male is the reference group for gender while female is the focal group. Similarly, Caucasian is the reference group for race/ethnicity with the minority race/ethnic groups being focal groups. Separate models are run for DIF for male versus female, black versus white, Hispanic versus white, Asian versus white, and Native American versus white.

Items are flagged for DIF using a blended approach that employs a likelihood ratio test of statistical significance to determine if DIF is present, and then assessing whether any evidence of DIF is practically significant using the Nagelkerke R^2 statistic (1991). For DIF investigations using LR, a common effect size measure is defined as the difference in R^2 values between the full model (including ability, subgroup membership, and their interaction as predictors) and the base model (including ability as the sole predictor). Items are categorized as exhibiting negligible DIF if the null hypothesis is not rejected or the difference in R^2 statistics is less than 0.035, moderate DIF if the null hypothesis is rejected and the difference in R^2 statistics is greater than or equal to 0.035 and less than 0.070, or large DIF if the null hypothesis is rejected and the difference in R^2 statistics is greater than or equal to 0.070 (Jodion & Gierl, 2001).

There are a couple of points in the Star Math assessment development cycle when items are evaluated for DIF. The first time point is when an item is included as a field test item as part of Star Math's item calibration process. During item calibration, new assessment items are tried out with different groups of students to make sure that items have appropriate statistical and psychometric properties before they are used operationally and count towards a student's score. The second time point is when the full item bank of operational test items is recalibrated for scale maintenance, to check whether the statistical and psychometric properties of the items have remained similar after the items become operational.

It is important to point out that just because an item is flagged for DIF against one or more subgroups does not necessarily mean that the item is biased. There are many possible explanations why an item may be statistically flagged for DIF. All items that are statistically flagged as having non-negligible DIF are marked for a bias and sensitivity review by the Content team. This review process consists of several subject matter experts with diverse perspectives and different backgrounds looking at and reviewing each item to see if there is any content in the item that may be biased against a particular subgroup and might explain why the item was statistically flagged for DIF. Items identified as being biased for any reason are removed from the item bank and do not appear on the Star Math test. The statistical flagging of items for DIF as well as the bias and sensitivity review by the Content team helps ensure test fairness and that the items that appear on Star Math do not favor any group of students that may take the test.

As shown in Figure 5, only 1% of over 5000 items in the Star Math item bank showed any evidence of DIF when Star Math was recalibrated in 2021.

Figure 5: Summary of Star Math Items with DIF

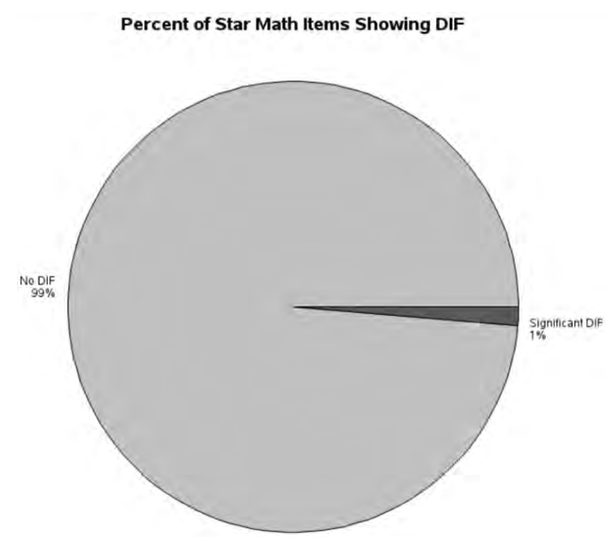


Table 34 shows the DIF results by reference and focal groups from various DIF analyses. These results suggest that of the thousands of items analyzed, very few items were flagged for DIF. There were 0.27% of items categorized with non-negligible DIF for male versus female, 0.00% of items flagged with non-negligible

DIF for Asian versus white, 0.27% of items flagged with non-negligible DIF for black versus white, 0.16% of items flagged with non-negligible DIF for Hispanic versus white, and 0.00% of items flagged with non-negligible DIF for Native American versus white. These results provide evidence of the fairness of the Star Math test for different demographic subgroups that take the assessment. As previously noted, any items that show DIF are removed from operational use.

Table 34: Percentage of Items Showing DIF for Different DIF Comparisons

DIF Comparison	Percent of Items Showing DIF
Female versus Male	0.27%
Asian versus White	0.00%
Black versus White	0.27%
Hispanic versus White	0.16%
Native American versus White	0.00%

Summary of Star Math Validity Evidence

The validity data presented in this technical manual includes evidence of Star Math's concurrent, predictive, and construct validity, as well as classification accuracy statistics; strong measures of association with math achievement levels on state and multi-state accountability assessments; and extensive evidence of its technical adequacy for screening and progress monitoring. Exploratory and confirmatory factor analyses provided evidence that Star Math measures a unidimensional construct, consistent with the assumption underlying its use of the Rasch item response model. The Meta-Analysis section showed the average uncorrected correlation between Star Math and all other math tests to be 0.758. (Many meta-analyses adjust the correlations for range restriction and attenuation to less than perfect reliability; had we done that here, the average correlation would have exceeded 0.80.) Correlations with specific measures of math ability were often higher than this average. For example, correlations with PARCC assessments averaged 0.83, and those with Smarter-Balanced Assessment scores averaged 0.88. The overall pattern of hundreds of correlations between Star Math and scores on other recognized math assessments provides strong support for the claim that Star Math is a measure of math achievement.

Finally, the data showing the relationship between the current, standards-based Star Math Enterprise test and scores on specific state accountability tests and on the SBAC and PARCC Common Core consortium tests show that the correlations with these important measures are consistent with the meta-analysis findings.

Norming

Two distinct kinds of norms are described in this chapter: test score norms and growth norms. The former refers to distributions of test scores themselves. The latter refers to distributions of changes in test scores over time; such changes are generally attributed to growth in the attribute that is measured by a test. Hence distributions of score changes over time may be called “growth norms.”

Background

National norms for Star Math were first developed in 2002 for Version 1 of the assessment, then updated in 2012 and 2017. The 2017 norms were used since the 2017–2018 school year until new norms were developed for introduction at the start of the 2022–2023 school year. This chapter describes the development of the 2022 norms.

The 2022 Star Math Norms

Prior to the development of the 2022 Star Math norms, a new reporting scale was developed, called the Unified scale. The Unified scale is a new linear transformation of the Star Math Rasch scores to a scale that shares features with a new scale developed for use with Star Reading and Star Early Literacy. The introduction of the Star Unified Scale provides a common scale that makes it possible for the first time to report performance on all Star assessments on the same scale.

The original Star Math scale is now referred to as the “Enterprise” score scale and will be available during the planned transition to the Unified scale as the default reporting scale. The Unified scale is the default scale in the software for the 2022–2023 school year. This chapter includes displays of normative summary data for both the Enterprise and the Unified scales

Due to testing impacts from COVID-19, the 2022 Star Math norms are based on Star Math test data collected over the course of the 2018–2019 school year. Separate early fall and late spring norms were developed for grades K through 12. Students participating in the norming study took assessments between August 1, 2018 and June 30, 2019. Students took the Star Math tests under normal test administration conditions. No specific norming test was developed and no deviations were made from the usual test administration. Thus, students in the norming sample took Star Math tests as they are administered in everyday use.

Sample Characteristics

During the norming period, a total of 4,990,729 US students in grades K–12 took Star Math tests administered using Renaissance servers hosted by Renaissance Learning. The first step in sampling was to select a representative sample of students who had tested in the fall, in the spring, or in both the fall and spring of the 2018–2019 school year. From the fall and the spring samples, stratified subsamples were randomly drawn based on student grade and ability decile. The grade and decile sampling was necessary to ensure adequate and similar numbers of students in each grade, and each decile within grade. Because these norming data were a convenience sample drawn from the Star Math customer base, steps were taken to ensure the resulting norms were nationally representative of grades K–12 US student population with regard to certain important characteristics. A post-stratification procedure was used to adjust the sample proportions to the approximate national proportions on three key variables: geographic region, district socio-economic status, and district/school size. These three variables were chosen because they had previously been used in Star Math norming studies to draw nationally representative samples, are known to be related to test scores, and were readily available for the schools in the Renaissance hosted database.

The final norming sample size, after selecting only students with test scores in either the fall or the spring or both fall and spring in the norming year and further sampling by grade and ability decile was 3,035,052 students in grades K–12. There were 2,225,100 students in the fall norming sample and 1,682,660 students in the spring norming sample; 872,708 students were included in both norming samples. These students came from 16,055 schools across 50 states and the District of Columbia.

Table 35 and Table 36 provide a breakdown of the number of students participating per grade in the fall and spring, respectively.

Table 35: Numbers of Students per Grade in the Fall Norms Sample

Grade	N	Grade	N	Grade	N
K	12,860	5	311,590	10	74,190
1	113,910	6	309,840	11	26,170
2	322,370	7	271,510	12	14,160
3	270,980	8	131,580	Total	2,225,100
4	308,380	9	57,560		

Table 36: Numbers of Students per Grade in the Spring Norms Sample

Grade	N	Grade	N	Grade	N
K	31,710	5	210,600	10	43,480
1	143,750	6	218,060	11	20,490
2	226,570	7	175,650	12	5,380
3	226,470	8	126,570	Total	1,682,660
4	213,980	9	39,950		

National estimates of US student population characteristics were obtained from two entities: the National Center for Educational Statistics (NCES) and Market Data Retrieval (MDR).

- National population estimates of students' demographics (ethnicity and gender) in grades K–12 were obtained from NCES; these estimates were from 2017–2018 for private schools and 2018–2019 for public schools, the most recent data available. National estimates of race/ethnicity were computed using the NCES data based on single race/ethnicity and also a multiple-race category. The NCES data reflect the most recent census data from the US census bureau.
- National estimates of school-related characteristics were obtained from May 2018 Market Data Retrieval (MDR) information. The MDR database contains the most recent data on schools, some of which may not be reflected in the NCES data.

Table 37 on page 84 shows national percentages of children in grades K–12 by region, school/district enrollment, district socio-economic status, location, and school type (public versus private) along with the corresponding percentages in the norming sample. MDR estimates of geographic region were based on the four broad areas identified by the National Educational Association as Northeastern, Midwestern, Southeastern, and Western regions. The specific states in each region are shown below.

Geographic Region

Using the categories established by the National Center for Education Statistics (NCES), students were grouped into four geographic regions as defined below: Northeast, Southeast, Midwest, and West.

Northeast

Connecticut, District of Columbia, Delaware, Massachusetts, Maryland, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont

Southeast

Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia

Midwest

Iowa, Illinois, Indiana, Kansas, Minnesota, Missouri, North Dakota, Nebraska, Ohio, South Dakota, Michigan, Wisconsin

West

Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, New Mexico, Nevada, Oklahoma, Oregon, Texas, Utah, Washington, Wyoming

School size

Based on total school enrollment, schools were classified into one of three school size groups: small schools had under 200 students enrolled, medium schools had 200–499 students enrolled, and large schools had 500 or more students enrolled.

Socioeconomic status as indexed by the percent of school students with free and reduced lunch

Schools were classified into one of four classifications based on the percentage of students in the school who had free or reduced student lunch. The classifications were coded as follows:

- High socioeconomic status (0%–24%)
- Above-median socioeconomic status (25%–49%)
- Below-median socioeconomic status (50%–74%)
- Low socioeconomic status (75%–100%)

No students were sampled from the schools that did not report the percent of school students with free and reduced lunch.

The norming sample also included private schools, Catholic schools, students with disabilities, and English Language Learners as described below.

Table 37: Sample Characteristics Along with National Population Estimates and Sample Estimates

		National Estimates	Fall Norming Sample	Spring Norming Sample
Region	Midwest	21.0%	22.3%	23.5%
	Northeast	18.6%	13.2%	14.0%
	Southeast	25.0%	28.2%	26.1%
	West	35.4%	36.4%	36.4%
School Enrollment	< 200	3.7%	4.2%	4.5%
	200–499	27.9%	36.6%	37.5%
	≥ 500	68.4%	59.2%	58.0%
District Socioeconomic Status	Low	20.7%	23.6%	23.9%
	Below Median	21.5%	24.0%	23.3%
	Above Median	24.4%	23.4%	23.0%
	High	33.5%	29.0%	29.8%
Location	Rural	14.4%	21.0%	20.3%
	Suburban	41.7%	36.9%	37.4%
	Town	11.4%	15.0%	14.9%
	Urban	32.5%	27.1%	27.3%
School Type	Public	91.9%	92.7%	91.6%
	Non-Public	8.1%	7.3%	8.4%

Table 38 provides information on the demographic characteristics of students in the sample and national percentages provided by NCES. No weighting was done on the basis of these demographic variables; they are provided to help describe the sample of students and the schools they attended. Because Star assessment users do not universally enter individual student demographic information such as gender and ethnicity/race, some students were missing demographic data, and the sample summaries in Table 38 are based on only those students that had gender and ethnicity information available. In addition to the student demographics shown, an estimated 6.9% of the students in the norming sample were gifted and talented (G&T) as approximated by the 2011–2012 school data collected by the Office of Civil Rights (OCR). OCR is a subsidiary of the US Department of Education.

School type was defined to be either public (including charter schools) or non-public (private, Catholic).

Table 38: Student Gender and School Information: National Estimates and Samples Percentages

			National Estimate	Fall Norming Sample	Spring Norming Sample
Gender	Public	Female	48.7%	49.5%	50.2%
		Male	51.3%	50.5%	49.8%
	Non-Public	Female	–	51.4%	52.0%
		Male	–	48.6%	48.0%
Race/Ethnicity	Public	American Indian	1.0%	1.6%	1.7%
		Asian	5.6%	5.1%	4.8%
		Black	15.1%	17.6%	18.2%
		Hispanic	27.1%	23.6%	23.9%
		White	47.1%	48.7%	48.1%
		Multiple Race[a]	4.0%	3.4%	3.4%
	Non-Public	American Indian	0.6%	0.6%	0.9%
		Asian	7.2%	10.7%	9.9%
		Black	9.2%	6.1%	7.1%
		Hispanic	11.5%	34.2%	35.0%
		White	66.7%	41.3%	40.6%
		Multiple Race[a]	4.9%	7.2%	6.6%

a. Students identified as belonging to two or more races.

Test Administration

All students took current version Star Math tests under normal administration procedures. Some students in the normative sample took the assessment two or more times within the norming windows; scores from their initial test administration in the fall and the last test administration in the spring were used for computing the norms.

Data Analysis

Student test records were compiled from the complete database of Star Math Renaissance users. Data spanned one school year from August 2018 to June 2019. Students’ Unified scale Rasch scores on their first Star Math test taken during the first month of the school year based on grade placement were used to compute norms for the fall; students’ Unified scale Rasch scores on the last

Star Math test taken during the 7th or 8th month of the school year were used to compute norms for the spring. Interpolation was used to estimate norms for times of the year between the first month in the fall and the last month in the spring. The norms were based on the distribution of Unified scale Rasch scores for each grade.

As noted above, a post-stratification procedure was used to approximate the national proportions on key characteristics. Post stratification weights from the regional, district socio-economic status, and school size strata were computed and applied to each student's unified Rasch ability estimate. Norms were developed based on the weighted Rasch ability estimates and then transformed to both Star Math Enterprise and Unified scaled scores. Table 39 provides descriptive statistics for each grade with respect to the normative sample performance, in the Unified scaled score units. Table 40 provides descriptive statistics for each grade with respect to the normative sample performance, in the Enterprise scaled score units.

Table 39: Descriptive Statistics for Weighted Scaled Scores by Grade for the 2018–2019 Norming Sample in the Unified Scale

Grade	Fall Unified Scaled Scores				Spring Unified Scaled Scores			
	N	Mean	Standard Deviation	Median	N	Mean	Standard Deviation	Median
K	12,860	725	50	717	31,710	797	53	799
1	113,910	774	55	771	143,750	852	55	851
2	322,370	862	53	864	226,570	919	51	919
3	270,980	915	52	917	226,470	971	56	975
4	308,380	967	57	971	213,980	1,009	59	1,013
5	311,590	1,009	61	1,012	210,600	1,044	60	1,048
6	309,840	1,049	64	1,055	218,060	1,078	66	1,083
7	271,510	1,072	69	1,080	175,650	1,094	73	1,098
8	131,580	1,092	75	1,099	126,570	1,111	77	1,116
9	57,560	1,097	74	1,105	39,950	1,114	76	1,117
10	74,190	1,099	76	1,106	43,480	1,116	78	1,118
11	26,170	1,116	76	1,118	20,490	1,124	76	1,127
12	14,160	1,124	77	1,127	5,380	1,129	76	1,133

Table 40: Descriptive Statistics for Weighted Scaled Scores by Grade for the 2018–2019 Norming Sample in the Enterprise Scale

Grade	Fall Enterprise Scaled Scores				Spring Enterprise Scaled Scores			
	N	Mean	Standard Deviation	Median	N	Mean	Standard Deviation	Median
K	12,860	192	82	179	31,710	309	86	312
1	113,910	271	90	267	143,750	398	90	397
2	322,370	415	86	418	226,570	508	83	507
3	270,980	501	85	504	226,470	592	91	598
4	308,380	586	93	593	213,980	655	95	661
5	311,590	653	99	658	210,600	711	98	717
6	309,840	719	104	728	218,060	766	107	774
7	271,510	757	112	770	175,650	792	119	799
8	131,580	789	122	801	126,570	820	125	827
9	57,560	797	120	809	39,950	824	123	829
10	74,190	800	124	811	43,480	828	127	831
11	26,170	828	123	831	20,490	840	124	845
12	14,160	841	125	846	5,380	849	124	855

Growth Norms

Student achievement typically is thought of in terms of status: a student's performance at one point in time. However, this ignores important information about a student's learning trajectory—how much students are growing over a period of time. When educators are able to consider growth information—the amount or rate of change over time—alongside current status, a richer picture of the student emerges, empowering educators to make better instructional decisions.

To facilitate deeper understanding of achievement, Renaissance Learning maintains growth norms for Star Assessments that provide insight both on growth to date and likely growth in the future. Growth norms are currently available for Star Math, Star Reading, and Star Early Literacy, and may be available for additional Star adaptive assessments in the coming years.

The growth model used by Star Assessments is Student Growth Percentile (SGP, Betebenner, 2009). SGPs were developed by Dr. Damian Betebenner, originally in partnership with several state departments of education.[1] It should be noted that the initial development of SGP involved annual state summative tests with reasonably constrained testing periods within each state. Because Star tests may be taken at multiple times throughout the year, a number of adaptations to the original model were made. For more information about Star Math SGPs, please refer to this overview: http://doc.renlearn.com/KMNet/R00571375CF86BBF.pdf

SGPs are norm-referenced estimates that compare a student's growth to that of his or her academic peers nationwide. Academic peers are defined as those students in the same grade with a similar score history. SGPs are generated via a process that uses quantile regression to provide a measure of how much a student changed from one Star testing window to the next relative to other students with similar score histories.

SGPs range from 1–99 and are interpreted similarly to Percentile Ranks, with 50 indicating typical or expected growth. For instance, an SGP score of 37 means that a student grew as much or more than 37 percent of her academic peers, and less than about 63 percent of her academic peers.

The Star Math SGP package also produces a range of future growth estimates. Those are mostly hidden from users but are presented in goal-setting and related applications to help users understand what typical or expected growth looks like for a given student. They are particularly useful for setting future goals and understanding the likelihood of reaching future benchmarks, such as likely achievement of proficient on an upcoming state summative assessment.

At present, the Star Math SGP growth norms are based on a sample of 15,216,272 matched student records from the 2016–2017, 2017–2018, and 2018–2019 school years across grades K–12. The sample included 6,140,587 unique students across all three school years. Table 41 provides a summary of the number of students and tests that were used when computing the SGP growth norms.

Table 41: Numbers of Students and Number of Tests Used in Computing SGP Growth Norms

Grade	Students	Tests	Grade	Students	Tests
K	75,312	133,011	7	602,124	1,484,164
1	554,506	1,397,878	8	571,915	1,372,359
2	678,687	1,788,294	9	309,539	643,840
3	719,905	1,879,416	10	240,530	486,311
4	730,753	1,910,923	11	164,052	313,490
5	737,422	1,913,347	12	105,295	181,071
6	679,115	1,712,168	Total	6,140,587[a]	15,216,272

a. This is the total number of unique students across *all* grades.

1. Core SGP documentation and source code are publicly available at https://cran.r-project.org/web/packages/SGP/index.html.

Score Definitions

Types of Test Scores

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

- *Scaled scores.* Star Math creates a virtually unlimited number of test forms as it dynamically interacts with the students taking the test. In order to make the results of all tests comparable, and in order to provide a basis for deriving the other types of test scores described below, it is necessary to convert the results of Star Math tests to scores on a common scale. Star Math software does this in two steps. First, maximum likelihood is used to estimate each student's score on the Rasch ability scale, based on the difficulty of the items administered, and the pattern of right and wrong answers. Second, the Rasch ability scores are converted to scaled scores. Two different score scales are now available in Star assessments: the original scaled scores, which are referred to as "Enterprise" scaled scores; and a new score, expressed on the "Unified" score scale, which was introduced with the 2017–2018 school year.

  Enterprise Scale Scores

  For Star Math, the "Enterprise" scale scores are the same scores that have been reported continuously since Star Math Version 1 was introduced in 1998. The range of reported Star Math Enterprise scores extends from 0 to 1400.

  Unified Scale Scores

  Renaissance developed a single score scale that applies to all Star assessments: the Unified score scale. Development began with equating each test's underlying Rasch ability scales to a common Rasch scale; the result was the "unified Rasch scale," which is an extension of the Rasch scale used in Star Reading. The next step was to develop an integer scale based on the unified Rasch scale, with new scale scores anchored to important points on the original Enterprise score scales of both tests. The end result was a reported score scale that extends from 200 to 1400.

  Star Math and Star Reading Unified reported scale scores range from 600 to 1400. Star Early Literacy Unified reported scale scores range from 200 to 1100. One benefit of the Unified scale is an improvement in certain properties of the scale scores: scores on both tests are much less variable from grade to grade; measurement error is likewise less variable; and Unified score reliability is slightly higher than that of the Enterprise scores.

- *Criterion-referenced scores* describe what a student knows or can do, relative to a specific content domain or to a standard. Such scores may be expressed either on a continuous score scale or as a classification. An example of a criterion-referenced score on a continuous scale is a percent-correct score, which expresses what proportion of test questions the student can answer correctly in the content domain. An example of a criterion-referenced classification is a proficiency category on a standards-based assessment: the student may be said to be "proficient" or not, depending on whether his score equals, exceeds, or falls below a specific criterion (the "standard") used to define "proficiency" on the standards-based test. The domain scores and mastery classification charts in the Diagnostic Report are criterion-referenced.
- *Norm-referenced scores* compare a student's test results to the results of other students who have taken the same test. In this case, scores provide a relative measure of student achievement compared to the performance of a group of students at a given time. Percentile Ranks and Grade Equivalents are the two primary norm-referenced scores provided by Star Math software. Both of these scores are based on a comparison of a student's test results to the data collected during the development of the 2022 Star Math norms.

Grade Equivalent (GE)

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

GE scores are often misinterpreted as though they convey information about what a student knows or can do—that is, as if they were criterion-referenced scores. To the contrary, GE scores are norm-referenced.

GEs in Star Math range from 0 to 12.9+, where 0 represents the beginning of grade K. Because Star Math norms go no lower than grade K, the GE for a score below the minimum for GE 0 will be reported as "< 0". The scale divides the academic year into 10 monthly increments, and is expressed as a decimal with the unit denoting the grade level and the individual "months" in tenths. Because Star Math norms are based on fall and spring score data only, monthly GE scores are derived through interpolation by fitting a curve to the grade-by-grade medians. Table 44, "Scaled Score to Grade Equivalent Conversions" on page 104 in the Conversion Tables chapter, contains the Star Math Scaled Score to GE conversions for both Unified and Enterprise scaled scores.

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

Grade Equivalent Cap

For customers who are using either Star Math or Star Math Enterprise on the Renaissance hosted platform, GE scores will be capped when they exceed three grade levels above the student's actual grade placement (seeTable 42). When a student's Scaled Score produces a GE that is greater than the start of three grades above the student's current grade, Star Math will report that student's GE is greater than the cap grade but will not report the specific GE score. Because this cannot happen to students in tenth grade or above, the potential for a capped GE will only exist for K–9 students. When applicable, the GE cap will now appear on all Star Math reports—even those showing test scores from tests taken prior to this update.

For example, a fourth grade student cannot receive a GE score above 7.0 at any time of the year. If their GE score is above a 7.0, the reports will show a capped GE score of "> 7".

Table 42: Grade Equivalents with GE Cap

Grade Placement	Grade Equivalent	Grade Equivalent Reported As
4.6	6.9	6.9
4.6	7.0	7.0
4.6	7.1	> 7

Comparing Star Math GEs with Those from Conventional Tests

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

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

Percentile Rank (PR)

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

The PR scale is not an equal-interval scale. For example, at grade placement of 7.0, Star Math Unified scaled scores of 1128 and 1149 correspond to PRs of 80 and 90, respectively. Thus, a difference of 21 scaled score points represents a 10-point difference in PR. However, at the same grade placement, scaled scores of 1080 and 1096 correspond to PRs of 50 and 60, respectively. In this case, a difference of 16 scaled points also represents a 10-point difference in PR. For this reason, PR scores should not be averaged or otherwise algebraically manipulated. NCE scores, described below, are much more appropriate for these types of calculations.

Table 45 on page 108 and Table 46 on page 112 in the Conversion Tables chapter contain abridged versions of both the Unified and the Enterprise Scaled Score to Percentile Rank conversion tables used by Star Math, which only shows the fall norms based on the first month of the school year by grade. The unabridged table includes norms for all the monthly grade placement values from 0.0 to 12.9. For each grade, the fall norms (grade placement ending in ".0") and spring norms (grade placement ending in ".9") were computed directly from data, while the intermediate norms were estimated by linearly interpolating between the two end points.

Normal Curve Equivalent (NCE)

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

Table 47 on page 115 in the Conversion Tables chapter lists the NCEs corresponding to integer PR values and facilitates the conversion of PRs to NCEs. Table 48 on page 116 provides the reverse conversions from NCE to PR. The NCE values are given as a range of scores that convert to the corresponding PR value.

Student Growth Percentile (SGP)

Student Growth Percentiles (SGPs) are a norm-referenced quantification of individual student growth derived using quantile regression techniques. An SGP compares a student's growth to that of his or her academic peers nationwide with a similar achievement history on Star assessments. Academic peers are students who

- are in the same grade,
- had the same scores on the current test and (up to) two prior tests from different testing windows, and
- took the most recent test and the first prior test on the same dates.

SGPs provide a measure of how a student changed from one Star testing window[1] to the next relative to other students with similar starting Star Math scores. SGPs range from 1–99 and interpretation is similar to that of Percentile Rank scores; lower numbers indicate lower relative growth and higher numbers show higher relative growth. For example, an SGP of 70 means that the student's growth from one test window to another exceeds the growth of 70% of students nationwide in

1. We collect data for our growth norms during three different time periods: fall, winter, and spring. More information about these time periods is provided on page 95..

the same grade with a similar Star Math score history. All students, no matter their starting Star score, have an equal chance to demonstrate growth at any of the 99 percentiles.

SGPs are often used to indicate whether a student's growth is more or less than can be expected. For example, without an SGP, a teacher would not know if a Scaled Score increase of 100 represents good, not-so-good, or average growth. This is because students of differing achievement levels in different grades grow at different rates relative to the Star Math scale. For example, a high-achieving second-grader grows at a different rate than a low-achieving second-grader.

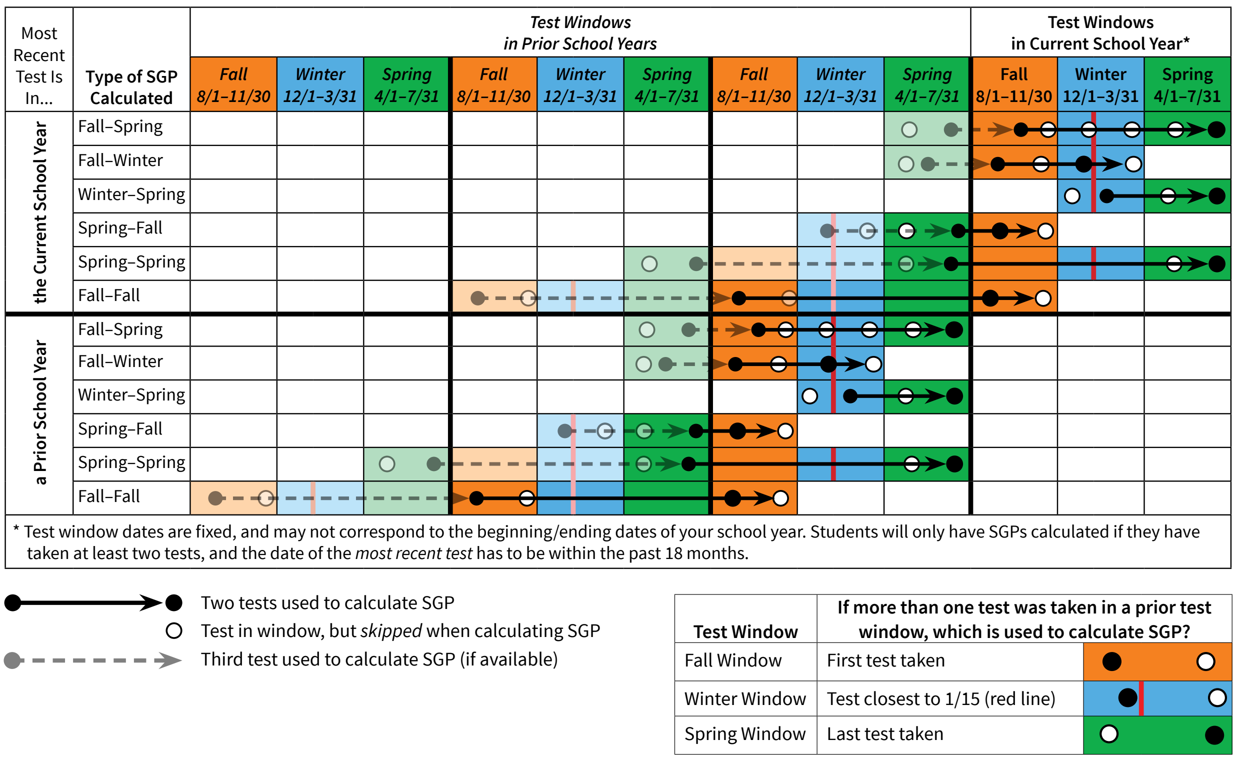
Similarly, a high-achieving second-grader grows at a different rate than a high-achieving eighth-grader.

SGPs can be aggregated to describe typical growth for groups of students—for example, a class, grade, or school as a whole—by calculating the group's median, or middle, growth percentile. No matter how SGPs are aggregated, whether at the class, grade, or school level, the statistic and its interpretation remain the same. For example, if the students in one class have a median SGP of 62, that particular group of students, on average, achieved higher growth than their academic peers.

SGP is calculated for students who have taken at least two tests (a current test and a prior test) within at least two different testing windows (Fall, Winter, or Spring).

If a student has taken more than one test in a single test window, the SGP calculation is based off the following tests:

- The current test is always the last test taken in a testing window.
- The test used as the prior test depends on what testing window it falls in:
  - Fall window: The first test taken in the Fall window is used.
  - Winter window: The test taken closest to January 15 in the Winter window is used.
  - Spring window: The last test taken in the Spring window is used.



Grade Placement

Star Math software uses students' grade placement values when determining norm-referenced scores. The values of PR (Percentile Rank) and NCE (Normal Curve Equivalent) are based not only on what Scaled Score the student achieved, but also on the grade placement of the student at the time of the test. For example, a second-grader in the first month with a Unified Scaled Score of 935 would have a PR of 92, while a third-grader in the first month with the same Unified Scaled Score would have a PR of 64.

Thus, it is crucial that student records indicate the proper grade and month within grade when students take a Star Math test, and that any testing in July or August reflects the proper understanding of how Star software deals with those months in determining grade placement.

Indicating the Appropriate Grade Placement

The numeric representation of a student's grade placement is based on the specific month in which he or she takes a test. Although teachers indicate a student's grade level or Math Instructional Level (MIL) using whole numbers, the Star Math software automatically adds fractional increments to that grade based on the month of the test. To determine the appropriate increment, Star Math considers the standard school year to run from September–June and assigns increment values of 0.0–0.9 to these months. The increment values for July and August depend on the school year setting:

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

Table 43 summarizes the increment values assigned to each month.

Table 43: Incremental Grade Placement Values per Month

Month	Decimal Increment	Month	Decimal Increment
July	0.0 or 0.99[a]	January	0.4
August	0.0 or 0.99[a]	February	0.5
September	0.0	March	0.6
October	0.1	April	0.7
November	0.2	May	0.8
December	0.3	June	0.9

a. Depends on the school year entered.

If your school follows the standard school calendar used in Star Math and you will not be testing in the summer, assigning the appropriate grade placements for your students is automatic.

However, if you are going to test students in July or August, whether it is for a summer program or because your normal calendar extends into these months, grade placements become an extremely important issue.

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

Instructions for specifying grade levels can be found at https://star-help.renaissance.com/hc/en-us/articles/11970114767131.

Compensating for Incorrect Grade Placements

Teachers cannot make retroactive corrections to a student’s grade placement by editing the grade assignments in a student’s record or by adjusting the increments for the summer months after students have tested. In other words, the Star Math software cannot go back in time and correct scores resulting from erroneous grade placement information. Thus, it is extremely important for the test administrator to make sure that the proper grade placement procedures are followed.

Quantile Measures

The Quantile Measure is an auxiliary scale developed by MetaMetrics for reporting math test performance. As described by Petersen, Kolen, and Hoover (1989, p. 222) auxiliary score scales can be used to “convey additional normative information, test-content information, and information that is jointly normative and content based.” One such auxiliary scale is The Quantile® Framework for Mathematics, which was developed to appropriately match students with materials at a level where the student has the background knowledge necessary to be ready for instruction on new mathematical skills and concepts. The Quantile Framework, and the Quantile scale, have been adopted by numerous states, and a number of standardized test publishers, for use as a common measure of math achievement.

In cooperation with MetaMetrics®, beginning in mid-2019, users of Star Math had the option of including Quantile measures on certain Star Math score reports, for students in grades 3 through 8. (In 2021, this was expanded to include students in grades 1–12). Quantile measures reported by Star Math will range from EM400Q to 1600Q. (The “Q” suffix identifies the score as a Quantile measure.

Where it appears, the “EM” prefix (“Emerging Mathematician”) indicates a score that is below 0 on the Quantile scale; such scores are typical of beginning math students.)

The Quantile Framework is described in detail in MetaMetrics (2015); an overview of it is available in MetaMetrics (2004). Research to link Star Math scores to the Quantile scale was conducted by MetaMetrics in the 2018–2019 school year, when approximately 2 million students in grades 1 through 12 took Star Math Enterprise tests, along with grade-appropriate MetaMetrics linking items previously calibrated on the Quantile scale. Details of the research study methodology and results, including scale linking particulars, are set out in a technical report (MetaMetrics, 2019).

Conversion Tables

Table 44: Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
0	600	722	0	187
0.1	723	729	188	198
0.2	730	735	199	208
0.3	736	742	209	220
0.4	743	749	221	231
0.5	750	756	232	242
0.6	757	763	243	254
0.7	764	770	255	265
0.8	771	777	266	277
0.9	778	784	278	288
1	785	791	289	299
1.1	792	798	300	311
1.2	799	804	312	320
1.3	805	811	321	332
1.4	812	818	333	343
1.5	819	825	344	355
1.6	826	831	356	364
1.7	832	838	365	376
1.8	839	844	377	385
1.9	845	851	386	397
2	852	857	398	407
2.1	858	864	408	418
2.2	865	870	419	428
2.3	871	877	429	439
2.4	878	883	440	449
2.5	884	889	450	459
2.6	890	895	460	468
2.7	896	901	469	478
2.8	902	907	479	488

Table 44: Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
2.9	908	913	489	498
3	914	919	499	507
3.1	920	925	508	517
3.2	926	931	518	527
3.3	932	936	528	535
3.4	937	942	536	545
3.5	943	947	546	553
3.6	948	953	554	563
3.7	954	958	564	571
3.8	959	963	572	579
3.9	964	968	580	587
4	969	973	588	595
4.1	974	978	596	603
4.2	979	983	604	611
4.3	984	988	612	620
4.4	989	992	621	626
4.5	993	997	627	634
4.6	998	1002	635	642
4.7	1003	1006	643	649
4.8	1007	1010	650	655
4.9	1011	1014	656	662
5	1015	1019	663	670
5.1	1020	1023	671	676
5.2	1024	1026	677	681
5.3	1027	1030	682	688
5.4	1031	1034	689	694
5.5	1035	1038	695	701
5.6	1039	1041	702	706
5.7	1042	1045	707	712
5.8	1046	1048	713	717
5.9	1049	1051	718	722
6	1052	1054	723	727

Table 44: Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
6.1	1055	1057	728	732
6.2	1058	1060	733	737
6.3	1061	1063	738	741
6.4	1064	1066	742	746
6.5	1067	1069	747	751
6.6	1070	1071	752	754
6.7	1072	1074	755	759
6.8	1075	1076	760	763
6.9	1077	1079	764	768
7	1080	1081	769	771
7.1	1082	1083	772	774
7.2	1084	1085	775	777
7.3	1086	1087	778	781
7.4	1088	1089	782	784
7.5	1090	1091	785	787
7.6	1092	1092	788	789
7.7	1093	1094	790	792
7.8	1095	1096	793	795
7.9	1097	1097	796	797
8	1098	1099	798	800
8.1	1100	1100	801	802
8.2	1101	1101	803	803
8.3	1102	1102	804	805
8.4	1103	1104	806	808
8.5	1105	1105	809	810
8.6	1106	1106	811	811
8.7	1107	1107	812	813
8.8	1108	1108	814	815
8.9	1109	1109	816	816
9	1110	1110	817	818
9.1	1111	1111	819	820
9.2	1112	1112	821	821

Table 44: Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
9.3	1113	1113	822	823
9.4	1114	1114	824	824
9.5	1115	1115	825	826
9.6	1116	1116	827	828
9.7	1117	1117	829	829
9.8	1118	1118	830	831
9.9	1119	1119	832	833
10	1120	1120	834	834
10.1	1121	1121	835	836
10.2	1122	1122	837	837
10.3	1123	1123	838	839
10.4	1124	1124	840	841
10.5	1125	1125	842	842
10.6	1126	1126	843	844
10.7	1127	1127	845	846
10.8	1128	1128	847	847
10.9	1129	1129	848	849
11	1130	1130	850	850
11.1	1131	1131	851	852
11.2	1132	1132	853	854
11.3	1133	1133	855	855
11.4	1134	1134	856	857
11.5	1135	1135	858	859
11.6	1136	1136	860	860
11.7	1137	1137	861	862
11.8	1138	1138	863	863
11.9	1139	1139	864	865
12	1140	1140	866	867
12.1	1141	1141	868	868
12.2	1142	1142	869	870
12.3	1143	1143	871	872
12.4	1144	1144	873	873

Table 44: Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
12.5	1145	1145	874	875
12.6	1146	1146	876	876
12.7	1147	1147	877	878
12.8	1148	1148	879	880
12.9	1149	1149	881	881
12.9+	1150	1400	882	1400

Table 45: Scaled Score to Percentile Ranks Conversion by Grade on the Unified Scale

PR	Grade (First Month of School Year)												
	K	1	2	3	4	5	6	7	8	9	10	11	12
1	600	600	600	600	600	600	600	600	600	600	600	600	600
2	644	671	743	792	825	856	896	913	923	930	929	945	946
3	652	678	756	805	840	876	916	930	942	945	943	964	965
4	655	683	764	815	852	889	928	943	953	957	954	978	980
5	660	686	771	822	863	901	938	952	963	967	964	992	996
6	662	690	776	829	871	911	946	960	972	977	973	1001	1005
7	664	693	781	835	879	917	952	968	979	987	982	1009	1014
8	668	695	785	841	885	923	958	974	986	994	989	1016	1022
9	669	698	789	846	890	929	963	981	992	999	996	1024	1029
10	670	700	793	849	895	933	968	986	997	1005	1002	1029	1036
11	672	704	797	852	900	937	972	990	1002	1009	1007	1032	1039
12	674	706	800	855	904	941	976	995	1007	1013	1011	1036	1042
13	677	708	803	858	908	944	980	998	1011	1018	1015	1039	1045
14	678	711	806	861	911	947	984	1002	1015	1022	1020	1042	1048
15	679	713	809	863	913	950	987	1005	1019	1026	1024	1045	1051
16	680	715	811	865	916	953	990	1008	1022	1030	1028	1048	1055
17	681	718	813	868	919	955	993	1012	1026	1033	1032	1051	1058
18	682	719	816	870	921	958	996	1014	1029	1037	1035	1054	1061
19	684	721	818	872	923	960	999	1017	1032	1040	1038	1058	1065
20	685	723	820	874	925	962	1001	1020	1035	1043	1042	1061	1068
21	686	725	822	876	927	965	1003	1023	1038	1045	1044	1063	1071
22	687	726	823	878	929	967	1005	1025	1041	1047	1047	1066	1074
23	688	728	825	880	931	969	1008	1028	1044	1050	1050	1068	1076
24	690	731	827	882	933	972	1010	1030	1046	1052	1052	1070	1079
25	691	732	829	883	934	974	1012	1032	1049	1055	1055	1073	1081
26	-	734	831	885	936	976	1014	1035	1051	1058	1058	1074	1083
27	692	735	833	887	937	978	1015	1037	1054	1060	1061	1077	1085

Table 45: Scaled Score to Percentile Ranks Conversion by Grade on the Unified Scale

PR	Grade (First Month of School Year)												
	K	1	2	3	4	5	6	7	8	9	10	11	12
28	693	737	834	888	939	980	1017	1039	1056	1063	1063	1079	1088
29	694	739	836	889	940	982	1019	1041	1058	1066	1066	1080	1090
30	696	740	838	891	942	984	1021	1043	1061	1068	1069	1083	1092
31	697	742	839	892	944	985	1023	1045	1063	1071	1071	1085	1094
32	698	744	841	894	945	987	1025	1047	1065	1073	1074	1088	1096
33	699	745	842	895	947	988	1027	1049	1067	1075	1076	1090	1098
34	700	747	844	897	948	990	1029	1051	1070	1077	1078	1092	1100
35	701	748	845	898	950	992	1031	1053	1072	1079	1080	1094	1102
36	702	750	846	899	951	993	1032	1055	1074	1081	1082	1095	1104
37	703	752	848	901	953	994	1034	1057	1076	1083	1084	1097	1106
38	704	753	849	902	954	996	1036	1059	1078	1085	1086	1099	1107
39	705	754	850	904	956	997	1037	1061	1080	1087	1088	1101	1109
40	706	755	852	905	957	998	1039	1062	1082	1089	1090	1103	1111
41	707	757	853	906	959	1000	1040	1064	1084	1091	1092	1104	1112
42	708	759	854	908	960	1001	1042	1066	1086	1093	1093	1106	1114
43	-	760	855	909	962	1002	1044	1068	1087	1095	1095	1107	1116
44	709	762	857	910	963	1004	1045	1070	1089	1096	1097	1109	1118
45	711	763	858	911	964	1005	1047	1072	1091	1098	1098	1111	1119
46	712	765	859	912	966	1006	1048	1073	1093	1099	1099	1112	1121
47	713	766	860	913	967	1008	1050	1075	1095	1100	1101	1114	1122
48	714	768	861	914	969	1009	1051	1077	1096	1102	1103	1115	1124
49	715	769	863	916	970	1010	1053	1079	1098	1103	1104	1117	1125
50	717	771	864	917	971	1012	1055	1080	1099	1105	1106	1118	1127
51	718	773	865	918	972	1013	1056	1082	1101	1106	1107	1120	1129
52	719	774	866	920	974	1015	1058	1084	1103	1107	1109	1121	1130
53	720	776	867	921	975	1016	1059	1085	1104	1109	1111	1123	1132
54	721	777	869	922	976	1018	1061	1087	1106	1110	1112	1124	1134
55	722	779	870	924	978	1019	1062	1088	1108	1112	1114	1125	1135
56	724	781	871	925	979	1020	1064	1090	1110	1113	1116	1127	1136
57	725	783	872	926	980	1022	1066	1091	1111	1114	1117	1129	1138
58	726	784	873	927	982	1023	1067	1093	1113	1116	1119	1130	1140
59	727	786	875	929	983	1024	1069	1095	1115	1117	1120	1132	1141
60	728	788	876	930	984	1026	1071	1096	1117	1119	1122	1134	1144
61	729	789	877	931	986	1028	1072	1097	1118	1121	1123	1135	1145
62	731	791	879	932	987	1029	1074	1099	1120	1122	1125	1137	1147
63	732	792	880	934	989	1031	1075	1100	1122	1124	1127	1140	1149
64	733	794	881	935	990	1032	1077	1102	1123	1125	1128	1141	1151
65	734	796	882	936	991	1034	1079	1103	1125	1127	1130	1143	1153
66	735	797	883	937	992	1035	1080	1105	1127	1129	1132	1144	1154

Table 45: Scaled Score to Percentile Ranks Conversion by Grade on the Unified Scale

PR	Grade (First Month of School Year)												
	K	1	2	3	4	5	6	7	8	9	10	11	12
67	737	798	885	938	994	1037	1082	1106	1128	1131	1133	1146	1156
68	738	800	886	939	995	1039	1084	1108	1130	1133	1135	1148	1158
69	739	801	888	941	996	1040	1085	1110	1132	1134	1136	1149	1159
70	741	803	889	942	997	1042	1087	1111	1134	1136	1138	1151	1161
71	743	805	891	944	999	1044	1088	1113	1135	1138	1140	1153	1163
72	745	807	892	945	1000	1045	1090	1115	1137	1140	1143	1155	1165
73	746	809	894	947	1002	1047	1092	1116	1139	1141	1144	1157	1167
74	748	810	895	948	1004	1049	1093	1118	1141	1143	1146	1159	1169
75	750	812	897	950	1005	1050	1095	1120	1143	1145	1148	1161	1171
76	752	814	899	952	1007	1052	1096	1121	1144	1146	1150	1163	1173
77	754	816	901	953	1008	1054	1097	1123	1146	1148	1152	1166	1175
78	756	818	902	955	1010	1055	1099	1124	1148	1150	1154	1169	1177
79	758	819	904	956	1012	1057	1100	1126	1149	1152	1156	1171	1181
80	760	821	906	958	1014	1059	1102	1128	1151	1154	1158	1173	1183
81	762	823	908	960	1016	1061	1104	1130	1153	1156	1159	1175	1185
82	765	825	909	962	1018	1063	1106	1132	1155	1157	1162	1177	1187
83	767	828	911	964	1020	1066	1108	1134	1157	1159	1164	1180	1190
84	770	831	913	966	1021	1068	1110	1136	1159	1161	1166	1182	1192
85	772	834	915	968	1023	1070	1112	1138	1161	1164	1169	1185	1196
86	775	837	917	970	1026	1072	1114	1141	1164	1167	1172	1189	1199
87	778	841	920	972	1028	1075	1116	1143	1167	1170	1175	1193	1203
88	781	844	922	975	1031	1078	1118	1145	1170	1174	1179	1198	1206
89	786	847	925	977	1033	1080	1120	1147	1174	1179	1183	1203	1211
90	790	850	928	980	1036	1083	1122	1149	1176	1181	1186	1207	1218
91	795	853	931	984	1039	1087	1125	1151	1178	1183	1188	1210	1222
92	800	856	935	987	1042	1090	1127	1154	1181	1186	1191	1214	1227
93	804	859	939	991	1046	1093	1130	1156	1184	1189	1195	1220	1232
94	810	863	943	995	1050	1096	1134	1159	1188	1192	1198	1226	1238
95	818	868	948	999	1054	1100	1137	1162	1193	1196	1203	1232	1246
96	827	875	954	1003	1059	1105	1142	1167	1201	1201	1208	1239	1252
97	839	882	962	1010	1066	1112	1147	1173	1212	1207	1216	1247	1260
98	854	893	972	1018	1077	1120	1154	1184	1227	1218	1230	1261	1274
99	884	910	987	1034	1092	1134	1169	1212	1250	1237	1253	1284	1295

Table 46: Scaled Score to Percentile Ranks Conversion by Grade on the Enterprise Scale

PR	Grade (First Month of School Year)												
	K	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	60	104	221	300	354	404	469	497	513	525	523	549	551
3	73	115	242	321	378	437	502	525	544	549	546	580	582
4	78	123	255	338	398	458	521	546	562	569	564	603	606
5	86	128	266	349	416	477	538	560	578	585	580	625	632
6	89	134	274	360	429	494	551	573	593	601	595	640	647
7	92	139	282	370	442	503	560	586	604	617	609	653	661
8	99	143	289	380	451	513	570	596	616	629	621	664	674
9	100	147	295	388	460	523	578	608	625	637	632	677	686
10	102	151	302	393	468	530	586	616	634	647	642	686	697
11	105	157	308	398	476	536	593	622	642	653	650	690	702
12	108	160	313	403	482	543	599	630	650	660	656	697	707
13	113	164	318	408	489	547	606	635	656	668	663	702	712
14	115	169	323	412	494	552	612	642	663	674	671	707	716
15	117	172	328	416	497	557	617	647	669	681	677	712	721
16	118	175	331	419	502	562	622	651	674	687	684	716	728
17	120	180	334	424	507	565	627	658	681	692	690	721	733
18	121	182	339	427	510	570	632	661	686	699	695	726	738
19	125	185	343	430	513	573	637	666	690	703	700	733	744
20	126	188	346	434	517	577	640	671	695	708	707	738	749
21	128	191	349	437	520	582	643	676	700	712	710	741	754
22	130	193	351	440	523	585	647	679	705	715	715	746	759
23	131	196	354	443	526	588	651	684	710	720	720	749	762
24	134	201	357	447	530	593	655	687	713	723	723	752	767
25	136	203	360	448	531	596	658	690	718	728	728	757	770
26	-	206	364	451	534	599	661	695	721	733	733	759	773
27	138	208	367	455	536	603	663	699	726	736	738	764	777
28	139	211	369	456	539	606	666	702	729	741	741	767	782
29	141	214	372	458	541	609	669	705	733	746	746	769	785
30	144	216	375	461	544	612	673	708	738	749	751	773	788
31	146	219	377	463	547	614	676	712	741	754	754	777	791
32	147	222	380	466	549	617	679	715	744	757	759	782	795
33	149	224	382	468	552	619	682	718	747	760	762	785	798
34	151	227	385	471	554	622	686	721	752	764	765	788	801

Table 46: Scaled Score to Percentile Ranks Conversion by Grade on the Enterprise Scale

PR	Grade (First Month of School Year)												
	K	1	2	3	4	5	6	7	8	9	10	11	12
35	152	229	386	473	557	625	689	725	755	767	769	791	804
36	154	232	388	474	559	627	690	728	759	770	772	793	808
37	156	235	391	477	562	629	694	731	762	773	775	796	811
38	157	237	393	479	564	632	697	734	765	777	778	799	812
39	159	238	395	482	567	634	699	738	769	780	782	803	816
40	160	240	398	484	569	635	702	739	772	783	785	806	819
41	162	243	399	486	572	638	703	742	775	786	788	808	821
42	164	247	401	489	573	640	707	746	778	790	790	811	824
43	-	248	403	490	577	642	710	749	780	793	793	812	827
44	165	251	406	492	578	645	712	752	783	795	796	816	830
45	169	253	408	494	580	647	715	755	786	798	798	819	832
46	170	256	409	495	583	648	716	757	790	799	799	821	835
47	172	258	411	497	585	651	720	760	793	801	803	824	837
48	173	261	412	499	588	653	721	764	795	804	806	825	840
49	175	263	416	502	590	655	725	767	798	806	808	829	842
50	178	266	417	503	591	658	728	769	799	809	811	830	845
51	180	269	419	505	593	660	729	772	803	811	812	834	848
52	182	271	421	508	596	663	733	775	806	812	816	835	850
53	183	274	422	510	598	664	734	777	808	816	819	838	853
54	185	276	425	512	599	668	738	780	811	817	821	840	856
55	186	279	427	515	603	669	739	782	814	821	824	842	858
56	190	282	429	517	604	671	742	785	817	822	827	845	860
57	191	286	430	518	606	674	746	786	819	824	829	848	863
58	193	287	432	520	609	676	747	790	822	827	832	850	866
59	195	291	435	523	611	677	751	793	825	829	834	853	868
60	196	294	437	525	612	681	754	795	829	832	837	856	873
61	198	295	438	526	616	684	755	796	830	835	838	858	874
62	201	299	442	528	617	686	759	799	834	837	842	861	877
63	203	300	443	531	621	689	760	801	837	840	845	866	881
64	204	304	445	533	622	690	764	804	838	842	847	868	884
65	206	307	447	534	624	694	767	806	842	845	850	871	887
66	208	308	448	536	625	695	769	809	845	848	853	873	889
67	211	310	451	538	629	699	772	811	847	851	855	876	892
68	212	313	453	539	630	702	775	814	850	855	858	879	895

Table 46: Scaled Score to Percentile Ranks Conversion by Grade on the Enterprise Scale

PR	Grade (First Month of School Year)												
	K	1	2	3	4	5	6	7	8	9	10	11	12
69	214	315	456	543	632	703	777	817	853	856	860	881	897
70	217	318	458	544	634	707	780	819	856	860	863	884	900
71	221	321	461	547	637	710	782	822	858	863	866	887	903
72	224	325	463	549	638	712	785	825	861	866	871	890	907
73	225	328	466	552	642	715	788	827	864	868	873	894	910
74	229	330	468	554	645	718	790	830	868	871	876	897	913
75	232	333	471	557	647	720	793	834	871	874	879	900	916
76	235	336	474	560	650	723	795	835	873	876	882	903	920
77	238	339	477	562	651	726	796	838	876	879	886	908	923
78	242	343	479	565	655	728	799	840	879	882	889	913	926
79	245	344	482	567	658	731	801	843	881	886	892	916	933
80	248	347	486	570	661	734	804	847	884	889	895	920	936
81	251	351	489	573	664	738	808	850	887	892	897	923	939
82	256	354	490	577	668	741	811	853	890	894	902	926	942
83	260	359	494	580	671	746	814	856	894	897	905	931	947
84	265	364	497	583	673	749	817	860	897	900	908	934	951
85	268	369	500	586	676	752	821	863	900	905	913	939	957
86	273	373	503	590	681	755	824	868	905	910	918	946	962
87	278	380	508	593	684	760	827	871	910	915	923	952	968
88	282	385	512	598	689	765	830	874	915	921	929	960	973
89	291	390	517	601	692	769	834	877	921	929	936	968	981
90	297	395	521	606	697	773	837	881	925	933	941	975	993
91	305	399	526	612	702	780	842	884	928	936	944	980	999
92	313	404	533	617	707	785	845	889	933	941	949	986	1007
93	320	409	539	624	713	790	850	892	938	946	955	996	1016
94	330	416	546	630	720	795	856	897	944	951	960	1006	1025
95	343	424	554	637	726	801	861	902	952	957	968	1016	1038
96	357	435	564	643	734	809	869	910	965	965	977	1027	1048
97	377	447	577	655	746	821	877	920	983	975	990	1040	1061
98	401	464	593	668	764	834	889	938	1007	993	1012	1063	1084
99	450	492	617	694	788	856	913	983	1045	1024	1050	1100	1118

Table 47: Percentile Rank to Normal Curve Equivalent Conversions

PR	NCE	PR	NCE	PR	NCE	PR	NCE
1	1.0	26	36.5	51	50.5	76	64.9
2	6.7	27	37.1	52	51.1	77	65.6
3	10.4	28	37.7	53	51.6	78	66.3
4	13.1	29	38.3	54	52.1	79	67.0
5	15.4	30	39.0	55	52.6	80	67.7
6	17.3	31	39.6	56	53.2	81	68.5
7	18.9	32	40.1	57	53.7	82	69.3
8	20.4	33	40.7	58	54.2	83	70.1
9	21.8	34	41.3	59	54.8	84	70.9
10	23.0	35	41.9	60	55.3	85	71.8
11	24.2	36	42.5	61	55.9	86	72.8
12	25.3	37	43.0	62	56.4	87	73.7
13	26.3	38	43.6	63	57.0	88	74.7
14	27.2	39	44.1	64	57.5	89	75.8
15	28.2	40	44.7	65	58.1	90	77.0
16	29.1	41	45.2	66	58.7	91	78.2
17	29.9	42	45.8	67	59.3	92	79.6
18	30.7	43	46.3	68	59.9	93	81.1
19	31.5	44	46.8	69	60.4	94	82.7
20	32.3	45	47.4	70	61.0	95	84.6
21	33.0	46	47.9	71	61.7	96	86.9
22	33.7	47	48.4	72	62.3	97	89.6
23	34.4	48	48.9	73	62.9	98	93.3
24	35.1	49	49.5	74	63.5	99	99.0
25	35.8	50	50.0	75	64.2		

Table 48: Normal Curve Equivalent to Percentile Rank Conversions

NCE Range Low–High	PR	NCE Range Low–High	PR	NCE Range Low–High	PR	NCE Range Low–High	PR
1.0–4.0	1	36.1–36.7	26	50.3–50.7	51	64.6–65.1	76
4.1–8.5	2	36.8–37.3	27	50.8–51.2	52	65.2–65.8	77
8.6–11.7	3	37.4–38.0	28	51.3–51.8	53	65.9–66.5	78
11.8–14.1	4	38.1–38.6	29	51.9–52.3	54	66.6–67.3	79
14.2–16.2	5	38.7–39.2	30	52.4–52.8	55	67.4–68.0	80
16.3–18.0	6	39.3–39.8	31	52.9–53.4	56	68.1–68.6	81
18.1–19.6	7	39.9–40.4	32	53.5–53.9	57	68.7–69.6	82
19.7–21.0	8	40.5–40.9	33	54.0–54.4	58	69.7–70.4	83
21.1–22.3	9	41.0–41.5	34	54.5–55.0	59	70.5–71.3	84
22.4–23.5	10	41.6–42.1	35	55.1–55.5	60	71.4–72.2	85
23.6–24.6	11	42.2–42.7	36	55.6–56.1	61	72.3–73.1	86
24.7–25.7	12	42.8–43.2	37	56.2–56.6	62	73.2–74.1	87
25.8–26.7	13	43.3–43.8	38	56.7–57.2	63	74.2–75.2	88
26.8–27.6	14	43.9–44.3	39	57.3–57.8	64	75.3–76.3	89
27.7–28.5	15	44.4–44.9	40	57.9–58.3	65	76.4–77.5	90
28.6–29.4	16	45.0–45.4	41	58.4–58.9	66	77.6–78.8	91
29.5–30.2	17	45.5–45.9	42	59.0–59.5	67	78.9–80.2	92
30.3–31.0	18	46.0–46.5	43	59.6–60.1	68	80.3–81.7	93
31.1–31.8	19	46.6–47.0	44	60.2–60.7	69	81.8–83.5	94
31.9–32.6	20	47.1–47.5	45	60.8–61.3	70	83.6–85.5	95
32.7–33.3	21	47.6–48.1	46	61.4–61.9	71	85.6–88.0	96
33.4–34.0	22	48.2–48.6	47	62.0–62.5	72	88.1–91.0	97
34.1–34.7	23	48.7–49.1	48	62.6–63.1	73	91.1–95.4	98
34.8–35.4	24	49.2–49.7	49	63.2–63.8	74	95.5–99.0	99
35.5–36.0	25	49.8–50.2	50	63.9–64.5	75		

Appendix A: Star Math Blueprint Skills

The content blueprint is broadly categorized into four domains: 1) Numbers & Operations, 2) Algebra, 3) Geometry & Measurement, 4) Data Analysis, Statistics, and Probability. However, to ensure appropriate distribution of items within an individual test, the assessment blueprint uses six content domains by treating Numbers and Operations as two separate domains, as well as treating Geometry and Measurement as two separate domains.

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations	Count with objects and numbers	N02	Count objects grouped in tens and ones
		N04	Determine one more than or one less than a given number across decades
		N42	Count on by ones from a number less than 100
		N43	Count back by ones from a number less than 20
		N45	Complete a skip pattern starting from a multiple of 2, 5, or 10
		N46	Count on by 100s from any number
		N56	Count objects to 20
		N57	Identify a number to 20 represented by a point on a number line
		N58	Determine one more than or one less than a given number
		N59	Count by 2s to 50 starting from a multiple of 2
		N60	Count objects grouped in tens and ones
		N82	Locate a number to 20 on a number line
		N95	Determine ten more than or ten less than a given number
		N96	Count by 5s or 10s to 100 starting from a multiple of 5 or 10, respectively
		NA1	Complete a sequence of numbers to 10
		NA4	Answer a question involving an ordinal number up to “tenth”
		NFY	Complete a skip pattern of 2 or 5 starting from any number
		NFZ	Complete a skip pattern of 10 starting from any number
	Identify odd and even numbers	N97	Identify odd and even numbers less than 100
	Identify, compare, and order fractions	AJB	Compare monomial numerical expressions using the properties of powers
		E5A	Estimate fractions of a whole
		N21	Identify a fraction equivalent to a given fraction
		N27	Locate a mixed number on a number line

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Identify, compare, and order fractions (continued)	N67	Determine a pictorial model of a fraction of a set of objects
		N68	Locate a fraction on a number line
		N69	Identify equivalent fractions using models
		N77	Identify a fraction represented by a point on a number line
		N78	Compare fractions using models
		N87	Determine a pictorial model of a fraction of a whole
		N88	Order fractions using models
		N91	Compare fractions with unlike denominators
		NB3	Order fractions with unlike denominators in ascending or descending order
		NG1	Compare fractions with like denominators
	Relate a decimal number to a percent	N0W	Convert a decimal number in thousandths to a percentage
		N30	Convert a percentage to its decimal equivalent
		NFT	Convert a decimal number to a percentage
	Relate a decimal to a fraction	AJ1	Compare expressions involving unlike forms of real numbers
		N22	Convert a fraction or mixed number in hundredths or thousandths to a decimal number
		N23	Convert a decimal number in hundredths or thousandths to a fraction
		N81	Compare numbers in decimal and fractional forms
		NB1	Determine the decimal number equivalent to a fraction model
		NB2	Determine the fraction equivalent to a decimal number model
	Relate place and value to a decimal number	N24	Relate a decimal number through ten-thousandths to its word form
		N25	Identify the place of a digit in a decimal number through hundredths
		N26	Estimate a decimal number from its position on a number line
		N29	Round a decimal number to a specified place through hundredths
		N50	Read a decimal number through the hundredths place
		N51	Locate a decimal number to tenths on a number line

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Relate place and value to a decimal number (continued)	N54	Represent a decimal number in expanded form using powers of ten
		N55	Determine the decimal number represented in expanded form using powers of ten
		N71	Identify a pictorial model of tenths or hundredths of a decimal number
		N79	Compare decimal numbers through the hundredths place
		N80	Compare decimal numbers of differing places to thousandths
		N89	Order decimal numbers through the hundredths place
		N92	Order numbers in decimal and fractional forms
		NB5	Order decimal numbers of differing places to thousandths in ascending or descending order
		NB7	Convert a number less than 1 to scientific notation
		NB8	Convert a number less than 1 from scientific notation to standard form
		NB9	Determine the decimal number from a pictorial model of tenths or hundredths
		NBA	Identify a decimal number to tenths represented by a point on a number line
	Relate place and value to a whole number	N03	Relate a whole number to the word form of the number to 100
		N06	Order whole numbers to 1,000 in ascending or descending order
		N07	Relate a 3-digit whole number to its word form
		N08	Identify the place of a digit in a 3-digit number
		N09	Represent a 3-digit whole number in expanded form
		N11	Order 4-digit whole numbers in ascending or descending order
		N12	Relate a 4- or 5-digit whole number to its word form
		N14	Represent a 4-digit whole number in expanded form
		N16	Order 4- to 6-digit whole numbers in ascending or descending order
		N17	Relate a 7- to 10-digit whole number to the word form of the number
		N18	Determine the value of a digit in a 6-digit number
		N19	Represent a 5-digit whole number in expanded form

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Relate place and value to a whole number (continued)	N37	Convert a whole number greater than 10 to scientific notation
		N48	Determine the value of a digit in a 4- or 5-digit whole number
		N49	Determine which digit is in a specified place in a 4- or 5-digit whole number
		N61	Compare whole numbers to 100 using words
		N62	Order whole numbers to 100 in ascending order
		N64	Determine the 3-digit number represented as hundreds, tens, and ones
		N70	Round a 4- to 6-digit whole number to a specified place
		N74	Represent a 2-digit number as tens and ones
		N76	Compare whole numbers to 1,000 using the symbols <, >, and =
		N83	Determine the value of a digit in a 2-digit number
		N84	Represent a 3-digit number as hundreds, tens, and ones
		N86	Determine the 4-digit whole number represented in thousands, hundreds, tens, and ones
		N98	Determine the 2-digit number represented as tens and ones
		NAB	Recognize equivalent forms of a 3-digit number using hundreds, tens, and ones
		NAE	Represent a 4-digit whole number as thousands, hundreds, tens, and ones
		NAF	Determine the 4- or 5-digit whole number represented in expanded form
		NG0	Compare 4- or 5-digit whole numbers using the symbols <, >, and =
		NKE	Determine the expanded form, written in powers of ten, of a whole number to 1,000,000
	Add and subtract fractions with like denominators	C22	Add fractions with like 1-digit denominators
		C23	Subtract fractions with like 1-digit denominators
		W22	WP: Add fractions with like denominators no greater than 10 and simplify the sum
		W23	WP: Subtract fractions with like denominators no greater than 10
		WCE	WP: Subtract fractions with like denominators no greater than 10 and simplify the difference

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Add and subtract fractions with like denominators (continued)	WX2	WP: Subtract fractions with like denominators and simplify the difference
		WX3	WP: Add mixed numbers with like denominators and simplify the sum
		WX4	WP: Subtract mixed numbers with like denominators and simplify the difference
		WXZ	WP: Add fractions with like denominators and simplify the sum
	Add and subtract fractions with unlike denominators	C24	Add fractions with unlike 1-digit denominators
		C25	Subtract fractions with unlike 1-digit denominators
		C28	Add mixed numbers with unlike denominators
		C29	Subtract mixed numbers with unlike denominators
		C57	Add fractions with unlike denominators that have factors in common and simplify the sum
		C76	Add fractions with unlike denominators that have no factors in common
		C77	Subtract fractions with unlike denominators that have factors in common and simplify the difference
		C78	Subtract fractions with unlike denominators that have no factors in common
		CA7	Add fractions with unlike denominators and do not simplify the sum
		E24	Estimate the sum of fractions with unlike 1-digit denominators
		E25	Estimate the difference between fractions with unlike 1-digit denominators
		E28	Estimate the sum of mixed numbers
		E29	Estimate the difference between mixed numbers with unlike denominators
		W24	WP: Add fractions with unlike 1-digit denominators
		W25	WP: Subtract fractions with unlike 1-digit denominators
		W28	WP: Add mixed numbers with unlike denominators
		W29	WP: Subtract mixed numbers with unlike denominators

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Add and subtract whole numbers with regrouping	C05	Add three 1-digit numbers
		C08	Add a 2-digit number and a 1- or 2-digit number with regrouping
		C09	Subtract a 1- or 2-digit number from a 2-digit number with one regrouping
		C11	Subtract a 2- or 3-digit number from a 3-digit number with two regroupings
		C18	Add four 1- to 4-digit whole numbers
		C19	Subtract two 2- to 6-digit whole numbers
		C47	Add 2- and 3-digit numbers with no more than one regrouping
		C49	Add 3- and 4-digit whole numbers with regrouping
		C50	Subtract 3- and 4-digit whole numbers with regrouping
		C69	Add two 3-digit numbers with one regrouping
		C70	Subtract a 1- or 2-digit number from a 3-digit number with one regrouping
		C71	Subtract a 3-digit number from a 3-digit number with one regrouping
		C88	Determine a number pair that totals 100
		CEL	Subtract a smaller number from a 3- or 4-digit whole number in expanded form
		W08	WP: Add a 2-digit number and a 1- or 2-digit number with regrouping
		W09	WP: Subtract a 1- or 2-digit number from a 2-digit number with one regrouping
		W18	WP: Add 3- and 4-digit whole numbers with regrouping
		W19	WP: Subtract 3- and 4-digit whole numbers with regrouping
	Add and subtract whole numbers without regrouping	A38	Determine the missing portion in a partially screened (hidden) collection of up to 10 objects
		C06	Add a 2-digit number and a 1-digit number without regrouping
		C07	Subtract a 1-digit number from a 2-digit number without regrouping
		C43	Know basic addition facts to 10 plus 10
		C44	Know basic subtraction facts to 20 minus 10
		C67	Add two 2-digit numbers without regrouping

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Add and subtract whole numbers without regrouping (continued)	C87	Subtract a 2-digit number from a 2-digit number without regrouping
		E08	Estimate the sum of two 2-digit numbers
		E09	Estimate the difference of whole numbers less than 100
		E41	Estimate a sum or difference of 2- to 4-digit whole numbers using any method
		E55	Estimate a sum or difference of whole numbers to 10,000 by rounding
		N05	Add or subtract zero to or from any number less than 100
		N99	Determine equivalent forms of a number, up to 10
		W03	WP: Use basic addition facts to solve problems
		W04	WP: Use basic subtraction facts to solve problems
		W06	WP: Add a 2-digit number and a 1-digit number without regrouping
		W7B	WP: Estimate a sum or difference of two 3- or 4-digit whole numbers using any method
		WXP	WP: Subtract a 1-digit number from a 2-digit number without regrouping
		WXQ	WP: Add two 2-digit numbers without regrouping
		WXR	WP: Subtract a 2-digit number from a 2-digit number without regrouping
		WXU	WP: Determine a basic addition-fact number sentence for a given situation
		WXV	WP: Determine a basic subtraction-fact number sentence for a given situation
		WXW	WP: Add two 3-digit numbers without regrouping
		WXY	WP: Subtract a 3-digit number from a 3-digit number without regrouping
	Add or Subtract Decimal Numbers	C33	Determine the sum of a whole number and a decimal number to hundredths
		C35	Subtract a decimal number from a whole number
		C51	Determine money amounts that total $10
		C79	Add decimal numbers and whole numbers
		C93	Subtract two decimal numbers of differing places to thousandths
		C98	Add two decimal numbers of differing places to thousandths

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Add or Subtract Decimal Numbers (continued)	CEB	Add or subtract cent amounts to or from whole dollar amounts
		CEC	Add dollars and cents to cents
		CED	Add dollars and cents to dollars
		CEE	Subtract cents from dollars and cents
		E32	Estimate the sum of two decimal numbers
		E33	Estimate the sum of a whole number and a decimal number
		E34	Estimate the difference of two decimal numbers
		E35	Estimate the difference of a whole number and a decimal number
		E44	Estimate the difference of two decimal numbers through thousandths and less than 1 by rounding to a specified place
		E45	Estimate the sum of two decimal numbers through thousandths and less than 1 by rounding to a specified place
		W33	WP: Determine the sum of a decimal number and a whole number
		W35	WP: Subtract a decimal number from a whole number
		W54	WP: Determine the amount of change from whole dollar amounts
		W94	WP: Add or subtract decimal numbers through thousandths
		W95	WP: Add or subtract a decimal number through thousandths and a whole number
		W96	WP: Estimate the sum or difference of two decimal numbers through thousandths using any method
	Convert between an improper fraction and a mixed number	N28	Convert an improper fraction to a mixed number
		N72	Convert a mixed number to an improper fraction
	Determine a square root	N31	Evaluate the positive square root of a perfect square
		N32	Determine an approximate square root of a number
		NBB	Determine the square root of a perfect-square fraction or decimal
		NBC	Determine the two closest integers to a given square root
		NBD	Approximate the location of a square root on a number line
		NFV	Determine both square roots of a perfect square

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Divide a whole number resulting in a decimal quotient	C58	Divide a whole number by a 1-digit whole number resulting in a decimal quotient through thousandths
		C59	Divide a whole number by a 2-digit whole number resulting in a decimal quotient through thousandths
		W50	WP: Divide a whole number by a 1- or 2-digit whole number resulting in a decimal quotient
	Divide whole numbers with a remainder in the quotient	C17	Divide a 2- or 3-digit whole number by a 1-digit whole number with a remainder in the quotient
		C55	Divide a multi-digit whole number by a 2-digit whole number, with a remainder and at least one zero in the quotient
		C56	Divide a multi-digit whole number by a 2-digit whole number and express the quotient as a mixed number
		W17	WP: Divide a 2- or 3-digit whole number by a 1-digit whole number with a remainder in the quotient
		W49	WP: Solve a 2-step problem involving whole numbers
		W57	WP: Divide a whole number and interpret the remainder
		W7C	WP: Divide a 3-digit whole number by a 1-digit whole number with a remainder in the quotient
	Divide Whole Numbers without a Remainder in the Quotient	AMQ	Recognize equivalent multiplication or division expressions involving basic facts
		C15	Divide a 2-digit whole number by a 1-digit whole number with no remainder in the quotient
		C21	Divide whole numbers with no remainder in the quotient
		C73	Know basic division facts to 100 ÷ 10
		CEG	Know basic division facts for 11 and 12
		CEH	Complete a multiplication and division fact family
		CEP	Divide a multi-digit whole number by 10 or 100 with no remainder
		E15	Estimate the quotient of a 2-digit whole number divided by a 1-digit whole number with no remainder in the quotient
		E21	Estimate a quotient using any method
		W15	WP: Divide a 2-digit whole number by a 1-digit whole number with no remainder in the quotient
		W21	WP: Divide whole numbers with no remainder in the quotient
		W2S	WP: Solve a 2-step whole number problem using more than one operation

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Divide Whole Numbers without a Remainder in the Quotient (continued)	W53	WP: Divide objects into equal groups by sharing
		W58	WP: Estimate a quotient using any method
		W66	WP: Divide using basic facts to 100 ÷ 10
		W90	WP: Divide a 3-digit whole number by a 1-digit whole number with no remainder in the quotient
	Evaluate a Numerical Expression	A49	Evaluate a numerical expression involving one or more exponents and multiple forms of rational numbers
		AA1	Simplify a monomial numerical expression involving the square root of a whole number
		AFM	Apply the product of powers property to a monomial numerical expression
		AFN	Apply the power of a power property to a monomial numerical expression
		AFP	Apply the quotient of powers property to monomial numerical expressions
		AG8	Multiply monomial numerical expressions involving radicals
		AG9	Divide monomial numerical expressions involving radicals
		AGT	Multiply a matrix by a scalar
		AGU	Add or subtract matrices
		AGV	Multiply matrices
		AGZ	Simplify an nth root
		AH1	Add or subtract complex numbers
		AH3	Simplify an expression involving a complex denominator
		AH9	Determine the logarithmic form of an exponential equation
		AHB	Evaluate a logarithm by converting it to exponential form
		AJ0	Evaluate a multi-step numerical expression involving absolute value
		AJE	Add and/or subtract numerical radical expressions
		AJF	Multiply a binomial numerical radical expression by a numerical radical expression
		AJG	Rationalize the denominator of a numerical radical expression
		AJR	Determine the determinant of a matrix
		AJV	Simplify an expression with a fractional exponent
		AJW	Add and subtract radical expressions

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Evaluate a Numerical Expression (continued)	AJY	Write an imaginary number in standard form
		AMT	Evaluate a numeric expression involving two operations
		AP3	Determine the inverse of a matrix
		AP4	Multiply complex numbers
		AP5	Determine the magnitude of a vector
		AP6	Add or subtract vectors component-wise
		AP7	Evaluate a linear combination of vectors
		N33	Evaluate the nth root of a whole number
		N34	Evaluate a whole number raised to a whole number power
		N35	Evaluate a whole number raised to a negative power
		N36	Evaluate a whole number raised to a fractional power
		N93	Evaluate a numerical expression of four or more operations, with parentheses, using order of operations
		N94	Evaluate a numerical expression involving integer exponents and/or integer bases
		NB6	Evaluate an integer raised to a whole number power
		NM6	Write a whole number raised to a whole number power as a product
	Find prime factors, common factors, and common multiples	N38	Identify the prime factors of a 2-digit number
		N39	Determine the greatest common factor of two whole numbers
		N40	Determine the least common multiple of two whole numbers
	Multiply and divide with decimals	C36	Multiply two decimal numbers
		C37	Divide decimal numbers
		C83	Multiply decimal numbers less than one in hundredths or thousandths
		C84	Divide a decimal number through thousandths by a 1- or 2-digit whole number where the quotient has 2–5 decimal places
		C85	Divide a 1- to 3-digit whole number by a decimal number to tenths where the quotient is a decimal number to thousandths
		C86	Divide a decimal number by a decimal number through thousandths, rounded quotient if needed

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Multiply and divide with decimals (continued)	C94	Multiply a decimal number through thousandths by 10, 100, or 1,000
		C99	Divide a decimal number by 10, 100, or 1,000
		C9A	Divide a 1- to 3-digit whole number by a decimal number to tenths where the quotient is a whole number
		C9B	Divide a 2- or 3-digit whole number by a decimal number to hundredths or thousandths, rounded quotient if needed
		C9F	Multiply a decimal number through thousandths by a whole number
		CA0	Multiply decimal numbers greater than one where the product has 2 or 3 decimal places
		W36	WP: Multiply two decimal numbers
		W37	WP: Divide a whole number by a decimal number
		W60	WP: Estimate the product of two decimals
		W80	WP: Multiply a decimal number through thousandths by a whole number
		W81	WP: Divide a decimal through thousandths by a decimal through thousandths, rounded quotient if needed
		W86	WP: Solve a multi-step problem involving decimal numbers
		W9B	WP: Divide a decimal number through thousandths by a 1- or 2-digit whole number
		W9C	WP: Divide a whole number by a decimal number through thousandths, rounded quotient if needed
		W9D	WP: Estimate the quotient of two decimals
		W9E	WP: Solve a 2-step problem involving decimals
	Multiply and divide with fractions	ABF	Determine the reciprocal of a positive whole number, a proper fraction, or an improper fraction
		AF5	Determine the reciprocal of a negative rational number
		C26	Multiply a fraction by a fraction
		C27	Divide a fraction by a fraction
		C30	Multiply mixed numbers
		C31	Divide mixed numbers
		C61	Multiply a mixed number by a fraction
		C80	Multiply a mixed number by a whole number
		C81	Divide a fraction by a whole number resulting in a fractional quotient

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Multiply and divide with fractions (continued)	C82	Divide a whole number by a fraction resulting in a fractional quotient
		W59	WP: Multiply or divide a fraction by a fraction
		W71	WP: Multiply or divide two mixed numbers or a mixed number and a fraction
		W99	WP: Solve a 2-step problem involving fractions
		WA9	WP: Solve a multi-step problem involving fractions or mixed numbers
	Multiply whole numbers	C14	Multiply a 2-digit whole number by a 1-digit whole number with no regrouping
		C16	Multiply a 2-digit whole number by a 1- or 2-digit whole number with regrouping
		C52	Multiply a 1- or 2-digit whole number by a multiple of 10, 100, or 1,000
		C53	Apply the distributive property to multiply a multi-digit number by a 1-digit number
		C54	Multiply a 3- or 4-digit whole number by a 1-digit whole number
		C72	Use a multiplication sentence to represent an area or an array model
		C74	Multiply a 2-digit whole number by a 2-digit whole number
		C91	Know basic multiplication facts to 10 × 10
		CE0	Know multiplication tables for 2, 5, and 10
		CEF	Know basic multiplication facts for 11 and 12
		CEJ	Multiply a 1-digit whole number by a multiple of 10 to 100
		CEM	Multiply a 3-digit whole number by a 2-digit whole number
		CEN	Multiply three 1- and 2-digit whole numbers
		E14	Estimate the product of a 2-digit number and a 1-digit number
		E20	Estimate the product of whole numbers using any method
		W14	WP: Multiply a 2-digit whole number by a 1-digit whole number without regrouping
		W16	WP: Multiply a 2-digit whole number by a 1- or 2-digit whole number
		W20	WP: Multiply whole numbers
		W46	WP: Multiply a multi-digit whole number by a 1-digit whole number

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Multiply whole numbers (continued)	W51	WP: Solve a multi-step problem involving whole numbers
		W65	WP: Multiply using basic facts to 10 × 10
		W8F	WP: Estimate a product of two whole numbers using any method
	Perform operations with integers	C62	Add integers
		C63	Subtract integers
		C64	WP: Add and subtract using integers
		C65	Multiply integers
		C66	Divide integers
		W87	WP: Multiply or divide integers
	Solve a problem involving percents	C97	Determine a percent of a number given a percent that is not a whole percent
		C9C	Determine the percent one number is of another number
		C9D	Determine a number given a part and a decimal percentage or a percentage more than 100%
		W38	WP: Determine the percent a whole number is of another whole number, with a result less than 100%
		W39	WP: Determine a percent of a whole number using percents less than 100
		W40	WP: Determine a whole number given a part and a percent
		W84	WP: Determine the result of applying a percent of decrease to a value
		W85	WP: Answer a question involving a fraction and a percent
		W8B	WP: Determine a given percent of a number
		W8C	WP: Determine the percent one number is of another number
		W8D	WP: Determine a number given a part and a decimal percentage or a percentage more than 100%
		WA6	WP: Determine the percent of decrease applied to a number
		WA7	WP: Determine the percent of increase applied to a number
		WA8	WP: Determine the result of applying a percent of increase to a value
		WB1	WP: Estimate a given percent of a number

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Numbers & Operations (continued)	Solve a proportion, rate, or ratio	C38	Determine the percent a whole number is of another whole number
		C39	Determine a given percent of a number
		C40	Determine a whole number given a part and a percent
		C41	Solve a proportion involving whole numbers
		C42	Determine if ratios are equivalent
		CJ2	Solve a proportion that generates a linear equation
		CJ4	Solve a proportion that generates a quadratic equation
		E38	Estimate the percent a whole number is of another whole number
		E39	Estimate a given percent of a number
		E40	Estimate a whole number given a part and a percent
		W41	WP: Solve a proportion
		W42	WP: Determine if ratios are equivalent
		W73	WP: Determine the whole, given part-to-part ratio and a part, where the whole is greater than 50
		W82	WP: Determine a unit rate with a whole number value
		W88	WP: Determine a part, given part-to-whole ratio and the whole, where the whole is greater than 50
		W89	WP: Determine a part, given part-to-whole ratio and a part, where the whole is greater than 50
		W8A	WP: Determine the whole, given part-to-whole ratio and a part, where the whole is greater than 50
		WA0	WP: Determine a part given a ratio and the whole where the whole is less than 50
		WA1	WP: Determine the whole given a ratio and a part where the whole is less than 50
		WA2	WP: Use a unit rate, with a whole number or whole cent value, to solve a problem
		WAA	WP: Determine a part, given part-to-part ratio and the whole, where the whole is greater than 50
		WAB	WP: Determine a part, given part-to-part ratio and a part, where the whole is greater than 50
		WAC	WP: Determine a unit rate
		WAD	WP: Use a unit rate to solve a problem

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra	Determine a linear equation	A02	Use a 1-variable, 1-step equation to represent a verbal statement
		A06	Determine an equation for a line given a graph
		A42	Use a 2-variable equation to construct an input-output table
		A46	Use a 2-variable equation to represent a relationship expressed in a table
		A53	Determine an equation of a line in slope-intercept form given the slope and y-intercept
		A83	Determine an equation for a line given the slope of the line and a point on the line that is not the y-intercept
		A84	Determine an equation of a line in point-slope or slope-intercept form given two points on the line
		A9C	Determine the slope-intercept form or the standard form of a linear equation
		AA5	Determine the table of values that represents a linear equation with rational coefficients in two variables
		AA6	Determine a linear equation in two variables that represents a table of values
		AFD	Determine an equation for a line that goes through a given point and is parallel or perpendicular to a given line
		AKX	WP: Determine a trigonometric function that represents a situation
		AM3	Represent a proportional relationship as a linear equation
		AN4	Use a table to represent a linear function
		AP0	WP: Determine an exponential function that represents a situation such as exponential growth or decay
		APG	Determine an equation of a line in standard form given the slope and y-intercept
		APH	Determine an equation of a line in standard form given two points on the line
		GKL	Determine an equation for a line parallel or perpendicular to a given graphed line
		W83	Use a 2-variable linear equation to represent a situation
		W8E	WP: Use a 1-variable equation with rational coefficients to represent a situation involving two operations
		WA3	Use a 2-variable equation to represent a situation involving a direct proportion

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra (continued)	Determine a linear equation (continued)	WAF	WP: Use a 1-variable 1-step equation to represent a situation
		WB2	WP: Use a 2-variable equation with rational coefficients to represent a situation
	Determine a system of linear equations	AP2	Represent a system of linear equations as a single matrix equation
		W74	WP: Determine a system of linear equations that represents a given situation
	Determine the operation given a situation	A30	WP: Determine the operation needed for a given situation
		ACB	Translate a verbal statement into an algebraic equation
		AMR	Determine the operation needed to make a number sentence true
		C90	Use a division sentence to represent objects divided into equal groups
		W67	WP: Determine a multiplication or division sentence for a given situation
	Evaluate an algebraic expression or function	A33	Evaluate a 2-variable expression, with two or three operations, using whole number substitution
		A36	Evaluate a 2-variable expression, with two or three operations, using integer substitution
		A50	Evaluate a function written in function notation for a given value
		AK1	Write a quadratic equation given its solutions
		ANT	Determine values of the inverse of a function using a table or a graph
		W72	WP: Evaluate a 1- or 2-variable expression or formula using whole numbers
	Graph a 1-variable inequality	A09	Relate a 1-variable inequality to its graph
	Graph on a coordinate plane	A08	Relate a graph to a 2-variable linear inequality
		A25	Relate a graph to an equation of a parabola
		A26	Relate a graph of an ellipse centered at the origin to its equation
		A48	Determine the graph of a 1-operation linear function
		A52	Determine the graph of a linear equation given in slope-intercept, point-slope, or standard form
		A91	Determine the graph of a given quadratic function
		AA0	Determine the graph of a line using given information

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra (continued)	Graph on a coordinate plane (continued)	AA7	Determine the graph of a 2-operation linear function
		AA8	Determine the slope of a line given its graph or a graph of a line with a given slope
		AAC	Use a table to represent the values from a first-quadrant graph
		AFE	Determine the graph of a 2-variable absolute value equation
		AFL	Determine the graph of the solution set of a system of linear inequalities in two variables
		AHG	Determine the graph of a circle given the equation in standard form
		AHJ	Determine the graph of a hyperbola given the equation in standard form
		AHL	Determine the graph of a vertically oriented parabola
		AHM	Determine the graph of a horizontally oriented parabola
		AHV	Determine the graph of a sine, cosine or tangent function
		AJ8	Determine a 2-variable linear inequality represented by a graph
		AJA	Determine the graph of a 1-variable absolute value inequality
		AJN	Graph the inverse of a linear function
		AK4	Relate a quadratic inequality in two variables to its graph
		AKE	Graph an ellipse
		ANN	Determine the graph of a piecewise-defined function
		ANP	Determine the component form of a vector represented on a graph
		ANQ	Relate a graph to a polynomial function given in factored form
		ANR	Identify a complex number represented as a vector on a coordinate plane
		ANS	Relate a graph to a square or cube root function
		GFS	Determine the ordered pair of a point in the first quadrant
		GFV	Determine the ordered pair of a point in any quadrant
		GM3	Determine the location of an ordered pair in any quadrant
		W79	WP: Answer a question using the graph of a quadratic function

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra (continued)	Identify characteristics of a linear equation or function	AMJ	Determine the slope of a line given a table of values
		A19	Determine the slope of a line given the coordinates of two points on the line
		A20	Determine the x- or y-intercept of a line given a 1-variable equation
		A9A	WP: Determine a reasonable domain or range for a function in a given situation
		A9E	Determine the slope of a line given an equation in point-slope or slope-intercept form
		AA9	Determine the x- or y-intercept of a line given its graph
		AF6	Determine if a relation is a function
		AF7	Determine if a function is linear or nonlinear
		AF8	Determine whether a graph or a table represents a linear or nonlinear function
		AJ2	Determine the independent or dependent variable in a given situation
		AJ3	Determine the domain or range of a function
		AJ4	Determine if a table or an equation represents a direct variation, an inverse variation, or neither
		AJK	Identify the domain or range of a radical function
		AJL	Determine the domain and range of a graphed function
		AKC	Determine the domain of a rational function
		AM5	Determine the effect of a change in the slope and/or y-intercept on the graph of a line
		AM8	Determine the result of a change in a or c on the graph of y=ax^2 + c
		AP8	Identify the vertex, axis of symmetry, or direction of the graph of a quadratic function
		AP9	Identify the end behavior, asymptotes, excluded values, or behavior near excluded values of a rational function
		APA	WP: Interpret an interest rate, rate of change, initial amount, frequency of compounding and other parameters of an exponential function
		APB	Determine if the inverse of a function is a function
		APC	Determine the equation of the inverse of a linear, rational root, or polynomial function

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra (continued)	Identify characteristics of a linear equation or function (continued)	APD	Determine the equation of a function resulting from a translation and/or scaling of a given function
		APE	Determine the x- or y-intercept of a line given a 2-variable equation
		APF	Determine the slope of a line given the graph of the line
		GG1	Determine if lines through points with given coordinates are parallel or perpendicular
		GG2	Determine the coordinates of a point through which a line must pass in order to be parallel or perpendicular to a given line
		W76	WP: Interpret the meaning of the slope of a graphed line
		WB3	WP: Interpret the meaning of the y-intercept of a graphed line
	Relate a rule to a pattern	A21	Determine the common difference in an arithmetic sequence
		A22	Find a specified term in an arithmetic sequence
		A29	Extend a number pattern involving addition
		A31	Identify a missing term in a multiplication or a division number pattern
		A32	Determine the variable expression with one operation for a table of paired numbers
		A39	Determine the rule for an addition or subtraction number pattern
		A40	Identify a missing figure in a growing pictorial or non-numeric pattern
		A44	Generate a table of paired numbers based on a rule
		A95	Extend a number pattern involving subtraction
		AA4	Determine a rule that relates two variables
		ACA	Determine the algebraic equation that describes a pattern represented by data in a table
		AKL	Find a specified term of an arithmetic sequence given the first term and the common difference
		AKM	Find a specified term of an arithmetic sequence
		AKN	Find a specified term of an arithmetic sequence given the formula for the nth term
		AKP	WP: Solve a problem that can be represented by an arithmetic sequence

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra (continued)	Relate a rule to a pattern (continued)	AKR	Find a specified term of a geometric sequence
		AKS	Find a specified term of a geometric sequence given the first three terms of the sequence
		AMS	Extend a number pattern
		ANH	Determine the explicit formula for an arithmetic sequence
		ANJ	Identify a given sequence as arithmetic, geometric, or neither
		ANK	Find a specified term of a binomial expression raised to a positive integer power
		ANL	WP: Solve a problem that can be represented by a geometric sequence
		ANM	WP: Solve a problem that can be represented by a finite geometric series
		GJZ	Use inductive reasoning to determine a rule
		W7E	WP: Generate a table of paired numbers based on a variable expression with one operation
		W97	WP: Determine the variable expression with one operation for a table of paired numbers
	Simplify an Algebraic Expression	A12	Add or subtract polynomial expressions
		A13	Multiply two binomials
		A18	Factor a common term from a binomial expression
		A55	Simplify a rational expression involving polynomial terms
		A56	Multiply rational expressions
		A57	Divide a polynomial expression by a monomial
		A58	Add or subtract two rational expressions with unlike polynomial denominators
		A61	Simplify an algebraic expression by combining like terms
		A87	Apply the product of powers property to a monomial algebraic expression
		A88	Apply the power of a power property to a monomial algebraic expression
		A89	Apply the power of a product property to a monomial algebraic expression
		A8A	Apply the quotient of powers property to monomial algebraic expressions
		A8B	Apply the power of a quotient property to monomial algebraic expressions

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra (continued)	Simplify an Algebraic Expression (continued)	A8E	Multiply two binomials of the form (ax +/– b)(cx +/– d)
		A8F	Factor the GCF from a polynomial expression
		A90	Factor trinomials that result in factors of the form (ax +/– b)(cx +/– d)
		A97	Multiply two monomial algebraic expressions
		AA2	Simplify a monomial algebraic radical expression
		AAE	Apply terminology related to polynomials
		AAF	Multiply two binomials of the form (x +/– a)(x +/– b)
		AFQ	Simplify a polynomial expression by combining like terms
		AFR	Multiply a polynomial by a monomial
		AFS	Multiply two binomials of the form (ax +/– by)(cx +/– dy)
		AFV	Multiply a trinomial by a binomial
		AFW	Factor trinomials that result in factors of the form (x +/– a)(x +/– b)
		AFX	Factor a trinomial that results in factors of the form (ax +/– by)(cx +/– dy)
		AFY	Factor the difference of two squares
		AFZ	Factor a perfect-square trinomial
		AGA	Multiply monomial algebraic radical expressions
		AGB	Divide monomial algebraic radical expressions
		AGF	Divide rational expressions
		AGG	Divide a polynomial expression by a binomial
		AGJ	Add or subtract two rational expressions with like denominators
		AGK	Add or subtract two rational expressions with unlike monomial denominators
		AGP	Determine the composition of two functions
		AGY	Represent an algebraic radical expression in exponential form
		AH0	Simplify an expression with rational exponents
		AH7	Factor a polynomial using long division
		AH8	Factor a polynomial by grouping
		AHA	Convert between a simple exponential equation and its corresponding logarithmic equation

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra (continued)	Simplify an Algebraic Expression (continued)	AJC	Apply properties of exponents to monomial algebraic expressions
		AJD	Factor a polynomial that has a GCF and two linear binomial factors
		AJH	Rationalize the denominator of an algebraic radical expression
		AJJ	Add or subtract algebraic radical expressions
		AK6	Factor a difference of squares
		AK7	Factor the sum or difference of 2 cubes
		AK8	Factor a polynomial into a binomial and trinomial
		ANU	Simplify a monomial algebraic expression that includes fractional exponents and/or nth roots
		ANV	Multiply or divide functions
		AP1	Identify equivalent logarithmic expressions using the properties of logarithms
	Solve a linear equation	A01	Determine a missing addend in a number sentence involving 2-digit numbers
		A04	Determine a solution to a 2-variable linear equation
		A28	Determine a missing addend in a basic addition-fact number sentence
		A37	Solve a proportion involving decimals
		A43	Solve a 2-step linear equation involving integers
		A45	Solve a 1-step equation involving whole numbers
		A47	Solve a 1-step linear equation involving integers
		A51	Solve a 1-variable linear equation with the variable on both sides
		A81	Determine a missing subtrahend in a basic subtraction-fact number sentence
		A98	Solve a 1-step equation involving rational numbers
		A99	Solve a 2-step equation involving rational numbers
		AAB	Rewrite an equation to solve for a specified variable
		AF9	Solve a 1-variable linear equation that requires simplification and has the variable on one side
		AFA	Solve a direct or inverse variation problem
		AMN	Determine the missing subtrahend in a number sentence involving 3-digit numbers

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra (continued)	Solve a linear equation (continued)	AMP	Determine the missing dividend or divisor in a number sentence involving basic facts
		W75	WP: Solve a problem involving a 1-variable, 2-step equation
		WXS	WP: Determine a missing addend in a basic addition-fact number sentence
		WXT	WP: Determine a missing subtrahend in a basic subtraction-fact number sentence
	Solve a Linear Inequality	A07	Determine the solution set of a 1-variable linear inequality
		A62	Determine the graph of the solutions to a 2-step linear inequality in one variable
		A9B	Solve a 1-variable linear inequality with the variable on both sides
		AAA	Solve a 2-step linear inequality in one variable
		ADC	Solve a 1-variable linear inequality with the variable on one side
		AFB	Solve a 1-variable compound inequality
		AJ6	Solve a 2-variable linear inequality for the dependent variable
		AJ7	Determine if an ordered pair is a solution to a 2-variable linear inequality
		WB4	WP: Solve a problem involving a 2-step linear inequality in one variable
	Solve a Nonlinear Equation	A15	Solve a quadratic equation using the square root rule
		A16	Solve a quadratic equation by factoring
		A17	Determine the term needed to complete the square in a quadratic equation
		A54	Solve a radical equation that leads to a quadratic equation
		A59	Solve a rational equation involving terms with monomial denominators
		A60	Solve a rational equation involving terms with polynomial denominators
		A85	Solve a 1-variable absolute value inequality
		A93	Solve a quadratic equation using the quadratic formula
		AA3	Solve a radical equation that leads to a linear equation
		AG1	Solve a quadratic equation by taking the square root
		AG2	Determine the solution(s) of an equation given in factored form

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Algebra (continued)	Solve a Nonlinear Equation (continued)	AG3	Use the discriminant to determine the number of real solutions
		AH5	Solve a quadratic equation with complex solutions
		AHC	Solve a logarithmic equation
		AJ5	Solve a 1-variable absolute value equation
		AK2	Solve a cubic equation
		AKD	Write the equation of a circle given its center and radius
		ANZ	Solve a problem involving the Pythagorean identity sin^2(theta) + cos^2(theta) = 1
		GGQ	Determine an equation of a circle
		GGR	Determine the radius, center, or diameter of a circle given an equation
	Solve a system of linear equations	A14	Solve a system of linear equations in two variables using any method
		AF1	Solve a number problem that can be represented by a linear system of equations
		AFJ	Determine the number of solutions to a system of linear equations
		AGX	Solve a problem involving matrices
		AJQ	Solve a system of three equations
Geometry & Measurement	Determine a missing figure in a pattern	A96	Identify a missing figure in a repeating pictorial or non-numeric pattern
		G01	Identify a missing figure in a geometric pattern
	Determine a missing measure or dimension of a shape	G02	Relate the radius to the diameter in a circle
		G22	Determine a missing angle measure in a triangle
		G23	Use the Pythagorean theorem to determine a length
		G27	Determine a missing dimension given two similar shapes
		GE4	Determine the midpoint of a line segment given the coordinates of the endpoints
		GE6	Determine the measure of an angle formed by parallel lines and one or more transversals given an angle measure
		GF6	Determine the measure of an angle or the sum of the angles in a polygon
		GF9	Determine a length using parallel lines and proportional parts

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Geometry & Measurement (continued)	Determine a missing measure or dimension of a shape (continued)	GFA	Determine a length using the properties of a 45-45-90 degree triangle or a 30-60-90 degree triangle
		GFB	Solve a problem involving the length of an arc
		GFC	Determine the length of a line segment, the measure of an angle, or the measure of an arc using a tangent to a circle
		GFD	Determine a length using a line segment tangent to a circle and the radius that intersects the tangent
		GFE	Determine the measure of an arc or an angle using the relationship between an inscribed angle and its intercepted arc
		GFG	Solve a problem involving the distance formula
		GFH	Solve a problem using inequalities in a triangle
		GFJ	Determine a length in a complex figure using the Pythagorean theorem
		GG3	Solve for the length of a side of a triangle using the Pythagorean theorem
		GG4	WP: Determine a length or an angle measure using triangle relationships
		GG5	Determine the length of a side or the measure of an angle in congruent triangles
		GG6	WP: Solve a problem using the properties of angles and/or sides of polygons
		GG8	Determine the length of a side in one of two similar polygons
		GG9	Determine the length of a side or the measure of an angle in similar triangles
		GGA	Determine a length given the perimeters of similar triangles or the lengths of corresponding interior line segments
		GGB	Determine a length in a triangle using a midsegment
		GGE	WP: Determine a length using similarity
		GGP	Determine the measure of an arc or a central angle using the relationship between the arc and the central angle
		GHC	Solve a problem involving the midpoint formula
		GHE	Determine a length or an angle measure using the segment addition postulate or the angle addition postulate
		GHF	Solve a problem involving a bisected angle or a bisected segment

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Geometry & Measurement (continued)	Determine a missing measure or dimension of a shape (continued)	GHJ	Determine the measure of an angle in a figure involving parallel and/or perpendicular lines
		GHL	Determine the measure of an angle using angle relationships and the sum of the interior angles in a triangle
		GHM	Determine a length in a triangle using a median
		GHP	Solve a problem involving a point on the bisector of an angle
		GHQ	Determine a length or an angle measure using general properties of parallelograms
		GHR	Determine a length or an angle measure using properties of squares, rectangles, or rhombi
		GHS	Determine a length or an angle measure using properties of kites
		GHT	Determine a length or an angle measure using properties of trapezoids
		GHU	Determine a length or an angle measure in a complex figure using properties of polygons
		GKA	Determine the effect of a change in dimensions on the perimeter or area of a shape
		GMY	Determine the distance between two points on a coordinate plane
		GN0	Determine the measure of an angle formed by parallel lines and one or more transversals given algebraic expressions
		GN1	Use triangle inequalities to determine a possible side length given the length of two sides
		GN2	Determine the measure of an angle or an arc using a tangent to a circle
		WB0	WP: Solve a problem involving similar shapes
		WB5	WP: Use the Pythagorean theorem to find a length or a distance
	Identify congruence and similarity of geometric shapes	GA3	Identify figures that are the same size and shape
		GA4	Compare common objects to basic shapes
		GA8	Determine lines of symmetry
		GB0	Determine the result of a reflection, rotation, or translation
		GE7	Identify a triangle congruence postulate that justifies a congruence statement
		GF7	Identify a triangle similarity postulate that justifies a similarity statement

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Geometry & Measurement (continued)	Identify congruence and similarity of geometric shapes (continued)	GF8	Identify similar triangles using triangle similarity postulates or theorems
		GFF	Identify congruent triangles using triangle congruence postulates or theorems
		GH8	Determine the coordinates of a preimage or an image given a reflection across a horizontal line, a vertical line, the line y = x, or the line y = –x
		GHA	Determine the coordinates of the image of a figure after two transformations of the same type
		GL0	Identify congruent shapes
		GL1	Identify mirror images
	Solve a problem involving the area of a shape	G06	Determine the area of a square
		G07	Determine the area of a rectangle given the length and width
		G08	Determine the area of a right triangle
		G09	Determine the area of a circle
		G24	Use a formula to determine the area of a triangle
		G25	Determine the area of a complex shape
		G33	Solve a problem given the area of a circle
		GAD	Determine the area of a polygon on a grid
		GAF	Determine the missing side length of a rectangle given a side length and the area
		GE5	Determine the area of a right triangle or a rectangle given the coordinates of the vertices of the figure
		GGS	Determine the area of a quadrilateral
		GGT	Determine a length given the area of a parallelogram
		GGU	Determine the area of a sector of a circle
		GGV	Determine the length of the radius or the diameter of a circle given the area of a sector
		GGW	WP: Determine a length or an area involving a sector of a circle
		GGX	Determine the measure of an arc or an angle given the area of a sector of a circle
		GJ3	Determine the area or circumference of a circle given an equation of the circle
		GKT	Determine the area of a shape composed of rectangles given a picture on a grid

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Geometry & Measurement (continued)	Solve a problem involving the area of a shape (continued)	GN3	Determine a length given the area of a kite or rhombus
		GN4	Determine a length given the area of a trapezoid
		W56	WP: Determine the area of a rectangle
		W69	WP: Determine the area of a triangle
		W70	WP: Determine a missing dimension given the area and another dimension
		W98	WP: Determine the area of a square or rectangle
	Solve a problem involving the perimeter of a shape	G03	Determine the perimeter of a square
		G04	WP: Determine the perimeter of a rectangle
		G05	Determine the perimeter of a triangle
		G26	Solve a problem involving the circumference of a circle
		GAB	Determine the perimeter of a rectangle given a picture showing length and width
		GAC	Determine the missing side length of a rectangle given a side length and the perimeter
		WA4	WP: Determine the perimeter or the area of a complex shape
	Solve a problem involving the surface area or volume of a solid	G10	Determine the volume of a rectangular prism
		G31	Determine the surface area of a rectangular prism
		G32	WP: Find the surface area of a rectangular prism
		G34	Determine the volume of a rectangular or a triangular prism
		GGY	Determine a length given the surface area of a right cylinder or a right prism that has a rectangle or a right triangle as a base
		GH0	Solve a problem involving the volume of a right pyramid or a right cone
		GH1	Determine the surface area of a sphere
		GH2	Determine the volume of a sphere or hemisphere
		GJP	Solve a problem involving the surface areas of similar solid figures
		W61	WP: Solve a problem involving the volume of a geometric solid
		W62	WP: Determine the surface area of a geometric solid
		W7F	WP: Determine the volume of a rectangular prism

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Geometry & Measurement (continued)	Use the vocabulary of geometry and measurement	G12	Identify rays
		G13	Identify line segments
		G14	Identify parallel lines
		G15	Identify intersecting line segments
		G16	Identify perpendicular lines
		G19	Identify perpendicular or parallel lines when given a transversal
		G21	Classify an obtuse angle or an acute angle given a picture
		G30	Classify an angle given its measure
		G37	Determine the common attributes in a set of geometric shapes
		GA1	Use basic terms to describe position
		GA2	Identify a circle, a triangle, a square, or a rectangle
		GA5	Identify a line of symmetry
		GA6	Identify a shape with given attributes
		GA7	Identify a common solid shape
		GFZ	Classify a right angle or a straight angle given a picture
		GH7	Relate the coordinates of a preimage or an image to a translation described using mapping notation
		GH9	Relate the coordinates of a preimage or an image to a dilation centered at the origin
		GHD	Identify a relationship between points, lines, and/or planes
		GHG	Identify angle relationships formed by multiple lines and transversals
		GHH	Identify parallel lines using angle relationships
		GJS	Determine the angle of rotational symmetry of a figure
		GK0	Use deductive reasoning to draw a valid conclusion from conditional statements
		GK1	Identify a statement or an example that disproves a conjecture
		GK2	Identify a valid biconditional statement
		GKE	Determine the number of faces, edges, or vertices in a 3-dimensional figure
		GKH	Identify a cross section of a 3-dimensional shape
		GKJ	Relate a net to a 3-dimensional shape

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Geometry & Measurement (continued)	Use the vocabulary of geometry and measurement (continued)	GKN	Identify the converse, inverse, or contrapositive of a statement
		GKV	Determine attributes of a triangle or a quadrilateral from a model
		GKW	Relate a model of a triangle or a quadrilateral to a list of attributes
		GKX	Identify a picture of a 3-dimensional shape
		GKY	Name a 3-dimensional shape from a picture
		GMZ	Identify a geometric construction given an illustration
		MA1	Compare objects using the vocabulary of measurement
	Calculate elapsed time	M17	Calculate elapsed time exceeding an hour with regrouping
		MDB	Calculate elapsed time within an hour, given two clocks, with regrouping
		W68	WP: Calculate elapsed time exceeding an hour with regrouping hours
	Determine a measurement	AKV	Convert between degree measure and radian measure
		AKY	Determine the value of an inverse sine, cosine, or tangent expression
		G17	Identify angle relationships formed by parallel lines cut by a transversal
		G18	Identify angle relationships formed by intersecting lines
		G20	Determine the measure of a vertical angle or a supplementary angle
		GGJ	Determine a sine, cosine, or tangent ratio in a right triangle
		M01	Convert between inches, feet, and yards
		M02	Estimate the height or length of a common object in customary units
		M04	Convert between customary units of capacity
		M05	Convert within metric units of mass, length, and capacity
		M06	Determine the approximate value of a unit converted between customary and metric measures
		M07	Identify an angle given its measure
		M08	Estimate the height of a common object in metric units
		M09	Measure length in centimeters
		M11	Convert a rate from one unit to another with a change in one unit

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Geometry & Measurement (continued)	Determine a measurement (continued)	M12	Convert a rate from one unit to another with a change in both units
		M18	WP: Determine a measure of length, weight or mass, or capacity or volume using proportional relationships
		MA9	Measure length in inches
		MAA	Read a thermometer in degrees Fahrenheit or Celsius
	Relate money to symbols, words, and amounts	C89	Determine cent amounts that total a dollar
		MA2	Identify a coin or the value of a coin
		MA4	Determine the value of groups of coins to $1.00
		N75	Translate between a dollar sign and a cent sign
		NAC	Convert money amounts in words to amounts in symbols
	Tell time	M10	Tell time to the minute
		M15	Tell time to the quarter hour
		M16	Tell time to 5-minute intervals
		MA5	Tell time to the hour and half hour
		MD9	Convert hours to minutes or minutes to seconds
Data Analysis, Statistics, and Probability	Determine a measure of central tendency	S07	Determine the mean of a set of whole number data
		S08	Determine the median of a set of data given a frequency table
		S14	Determine the median of an odd number of data values
		SD3	Determine the median of an even number of data values
	Determine the probability of one or more events	S11	Determine the probability of a single event
		S12	Determine the probability of independent events
	Read or answer a question about charts, tables, or graphs	AME	Determine if a scatter plot shows a positive relationship, a negative relationship, or no relationship between the variables
		AMF	Make a prediction based on a scatter plot
		S00	Read a simple pictograph
		S01	Read a table
		S02	Read a bar graph
		S03	Read a circle graph
		S04	Answer a question using information from a table
		S05	Answer a question using information from a bar graph
		S06	Answer a question using information from a circle graph

Table 49: Star Math Blueprint Skills

Blueprint Domain	Blueprint Skillset	Skill Code	Star Math US Blueprint Skill
Data Analysis, Statistics, and Probability (continued)	Read or answer a question about charts, tables, or graphs (continued)	S13	Answer a question using information from a line graph
		S18	Answer a question using information from a pictograph (1 symbol = more than 1 object)
		S19	Answer a question using information from a bar graph with a y-axis scale by 2s
		S21	Read a double-bar graph
		S22	Answer a question using information from a double-bar graph
		S23	Answer a question using information from a circle graph using percentage calculations
		S24	Answer a question using information from a histogram
		SA1	Read a tally chart
		SA2	Read a line graph
		SD7	Read a 2-category tally chart
		SD9	Answer a question using information from a 2-category tally chart
		SDC	Read a line plot
		SDD	Answer a question using information from a line plot
		SE6	Answer a question using information from a scatter plot
	Use a chart, table, or graph to represent data	S15	Use a circle graph to represent percentage data
		S16	Use a histogram to represent data
		S17	Use a pictograph to represent data (1 symbol = more than 1 object)
		S20	Use a line graph to represent data
		S26	Use a bar graph with a y-axis scale by 2s to represent data
		SA3	Use a double-bar graph to represent data
		SD1	Use a line plot to represent data
		SD5	Use a scatter plot to organize data
		SD8	Use a 2-category tally chart to represent groups of objects (1 symbol = 1 object)
	Use a proportion to make an estimate	S25	Use a proportion to make an estimate, related to a population, based on a sample

Appendix B: Additional Evidence of Star Math Validity

The Validity chapter of this technical manual places its emphasis on summaries of Star Math validity evidence, and on recent evidence which comes primarily from the 34-item, standards-based version of the assessment introduced in 2011. However, the abundance of earlier evidence, and evidence related to the 24-item Star Math versions, is all part of the accumulation of technical support for the validity and usefulness of Star Math. Much of that cumulative evidence is presented in this appendix to ensure that the historical continuity of research in support of Star Math validity is not lost. The material that follows touches on the following list of topics:

- Relationship of Star Math Scores to Scores on Other Tests of Math Achievement
- Relationship of Star Math Scores to Teacher Ratings
- Linking Star and State Assessments: Comparing Student- and School-Level Data
- Classification Accuracy and Screening Data Reported to The National Center on Response to Intervention (NCRTI)

Relationship of Star Math Scores to Scores on Other Tests of Math Achievement

The technical manual for the earliest version of Star Math listed correlations between scores on that test and those on a number of other standardized measures of math achievement, obtained in 1998 for more than 9,000 students who participated in Star Math norming for that version of the program. The standardized tests included a variety of well-established instruments, including the California Achievement Test (CAT), the Comprehensive Test of Basic Skills (CTBS), the Iowa Tests of Basic Skills (ITBS), the Metropolitan Achievement Test (MAT), the Stanford Achievement Test (SAT), and several statewide tests. During the development of Star Math Version 2, additional correlations with external tests were obtained from a total of more than 8,000 tests administered in 2000 and 2001.

During the 2014 norming of Star Math, scores on other standardized tests were obtained for more than 30,000 additional students. All of the standardized tests listed above were included, plus others such as Northwest Evaluation Association (NWEA) and TerraNova. Scores on state assessments from the following states were also included: Arkansas, Connecticut, Delaware, Florida, Georgia, Kentucky, Idaho, Indiana, Illinois, Maryland, Michigan, Minnesota, Mississippi, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Dakota, Texas, Virginia, and Washington. The extent that the Star Math test correlates with these tests provides support for its construct validity. That is, strong and positive correlations between Star Math and these other instruments provide support for the claim that Star Math effectively measures mathematics achievement.

Table 50 and Table 51 present the correlational data from the 2000–2001 development of Star Math 2. Table 50 lists the correlational details for 4,996 students in grades 1–6; Table 51 lists counterpart data for 3,066 students in grades 7–12.

Table 52 through Table 55 present the correlation coefficients between the scores on the Star Math test and other test instruments subsequent to the Star Math 2 development in years ranging from 2002 through 2016. Table 52 and Table 53 display "concurrent validity" data, that is, correlations between Star Math norming study test scores and other tests administered within a two-month time period. Tests listed in Table 52 and Table 53 were administered between the fall of 2001 and the spring of 2013.

Table 54 and Table 55 display predictive validity data from the same period. Predictive validity provides an estimate of the extent to which scores on the Star Math test predicted scores on criterion measures given at a later point in time, operationally defined as more than 2 months between the Star test (predictor) and the criterion test. It provides an estimate of the linear relationship between Star scores and scores on measures covering a similar academic domain.

Table 50: External Validity Data—Star Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 1–6[a]

Test	Version	Date	Score	1		2		3		4		5		6	
				n	r	n	r	n	r	n	r	n	r	n	r
Achievement Level (RIT) Test															
RIT		F 01	SS	–	–	–	–	–	–	–	–	–	–	150	0.69*
California Achievement Test															
CAT	5th Ed.	S 01	SS	–	–	–	–	46	0.52*	–	–	–	–	–	–
Cognitive Abilities Test															
CogAT		F 00	SS	–	–	–	–	41	0.61*	–	–	–	–	–	–
CogAT		F 01	SS	–	–	45	0.73*	–	–	–	–	–	–	–	–

Table 50: External Validity Data—Star Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 1–6[a]

				1		2		3		4		5		6	
Test	Version	Date	Score	n	r	n	r	n	r	n	r	n	r	n	r
Comprehensive Test of Basic Skills															
CTBS	4th Ed.	S 01	GE	–	–	–	–	–	–	43	0.67*	–	–	–	–
CTBS	A-13	S 00	NCE	–	–	–	–	–	–	65	0.60*	–	–	–	–
CTBS	A-13	S 00	SS	–	–	–	–	–	–	–	–	44	0.70*	–	–
CTBS	A-13	S 01	GE	–	–	–	–	–	–	–	–	–	–	56	0.69*
CTBS	A-13	S 01	NCE	–	–	–	–	–	–	–	–	67	0.72*	–	–
CTBS	A-13	S 01	SS	–	–	–	–	–	–	42	0.61*	–	–	–	–
Connecticut Mastery Test															
Conn	2nd	F 00	SS	–	–	–	–	–	–	–	–	35	0.51*	–	–
Conn	3rd	F 01	SS	–	–	–	–	–	–	42	0.64*	–	–	27	0.52*
Des Moines Public School (Grade 2 pretest)															
DMPS		F 01	NCE	–	–	25	0.76*	–	–	–	–	–	–	–	–
Educational Development Series															
EDS	13C	S 01	GE	–	–	–	–	30	0.69*	–	–	–	–	–	–
EDS	14C	S 00	GE	–	–	–	–	–	–	32	0.44*	–	–	–	–
EDS	15C	F 01	GE	–	–	–	–	–	–	–	–	37	0.68*	–	–
Florida Comprehensive Assessment Test															
FCAT		S 01	NCE	–	–	–	–	–	–	–	–	73	0.65*	–	–
Iowa Tests of Basic Skills															
ITBS	Form A	S 01	NCE	–	–	–	–	73	0.45*	78	0.65*	–	–	–	–
ITBS	Form A	F 01	NCE	–	–	–	–	25	0.41*	25	0.35	23	0.33	86	0.81*
ITBS	Form A	F 01	SS	–	–	–	–	–	–	–	–	–	–	73	0.64*
ITBS	Form K	F 00	SS	–	–	–	–	–	–	–	–	–	–	20	0.92*
ITBS	Form K	S 01	NCE	–	–	101	0.67*	74	0.64*	31	0.25	11	0.58	31	0.62*
ITBS	Form K	F 01	NCE	–	–	–	–	10	0.78*	16	0.78*	9	0.54	18	0.63*
ITBS	Form K	F 01	SS	–	–	–	–	–	–	–	–	75	0.77*	68	0.71*
ITBS	Form L	S 01	NCE	–	–	–	–	13	0.5	46	0.81*	13	0.73*	–	–
ITBS	Form L	S 01	SS	–	–	–	–	–	–	11	0.81*	–	–	–	–
ITBS	Form L	F 01	NCE	–	–	–	–	–	–	–	–	69	0.66*	–	–
ITBS	Form M	S 99	NCE	–	–	–	–	–	–	–	–	–	–	19	0.68*
ITBS	Form M	S 00	NCE	–	–	–	–	–	–	–	–	28	0.65*	–	–

Table 50: External Validity Data—Star Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 1–6[a]

Test	Version	Date	Score	1		2		3		4		5		6	
				n	r	n	r	n	r	n	r	n	r	n	r
Iowa Tests of Basic Skills (continued)															
ITBS	Form M	S 01	NCE	–	–	19	0.81*	–	–	43	0.78*	–	–	–	–
ITBS	Form M	S 01	SS	–	–	–	–	47	0.39*	32	0.55*	–	–	–	–
ITBS	Form M	F 01	NCE	5	0.88*	–	–	–	–	15	0.82*	–	–	–	–
McGraw Hill Mississippi/Criterion Referenced															
McGraw		S 01	SS	–	–	–	–	–	–	–	–	121	0.52*	–	–
Metropolitan Achievement Test															
MAT	7th Ed.	F 01	NCE	–	–	–	–	–	–	–	–	–	–	15	0.84*
Michigan Education Assessment Program															
MEAP		S 01	SS	–	–	–	–	–	–	–	–	88	0.72*	–	–
Multiple Assessment Series (Primary Grades)															
Multiple		S 01	NCE	–	–	14	0.52	19	0.54*	–	–	–	–	–	–
New York State Math Assessment															
NYSMA		S 01	SS	–	–	–	–	–	–	–	–	50	0.79*	–	–
North Carolina End of Grade															
NCEOG		F 01	SS	–	–	–	–	85	0.57*	–	–	–	–	–	–
Northwest Evaluation Association Levels Test															
NWEA		S 01	NCE	–	–	–	–	–	–	–	–	83	0.81*	64	0.78*
NWEA		F 01	NCE	–	–	–	–	50	0.56*	49	0.54*	99	0.70*	–	–
Ohio Proficiency Test															
Ohio		S 01	SS	–	–	–	–	113	0.65*	–	–	–	–	–	–
Stanford Achievement Test															
SAT9		S 99	SS	–	–	–	–	–	–	–	–	55	0.65*	–	–
SAT9		S 00	SS	–	–	–	–	–	–	–	–	–	–	15	0.5
SAT9		F 00	NCE	–	–	–	–	17	0.84*	20	0.83*	–	–	–	–
SAT9		F 00	SS	–	–	–	–	–	–	–	–	–	–	46	0.58*
SAT9		S 01	NCE	–	–	–	–	43	0.69*	–	–	50	0.38*	–	–
SAT9		S 01	SS	64	0.52*	–	–	–	–	58	0.41*	52	0.58*	51	0.65*
SAT9		F 01	SS	–	–	–	–	–	–	90	0.54*	32	0.67*	24	0.57*
Tennessee Comprehensive Assessment Program, 2001															
TCAP	2001	S 01	SS	–	–	–	–	–	–	–	–	48	0.56*	–	–

Table 50: External Validity Data—Star Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 1–6[a]

Test	Version	Date	Score	1		2		3		4		5		6	
				n	r	n	r	n	r	n	r	n	r	n	r
TerraNova															
TerraNova		S 00	NCE	–	–	–	–	–	–	–	–	–	–	43	0.60*
TerraNova		S 00	SS	–	–	–	–	–	–	–	–	11	0.61*	–	–
TerraNova		F 00	SS	–	–	–	–	–	–	–	–	108	0.62*	–	–
TerraNova		S 01	NCE	–	–	–	–	–	–	–	–	69	0.40*	85	0.62*
TerraNova		S 01	SS	–	–	–	–	–	–	104	0.50*	62	0.59*	131	0.71*
TerraNova		F 01	NCE	–	–	58	0.38*	63	0.56*	70	0.74*	85	0.61*	–	–
Test of New York State Standards															
TONYSS		S 01	SS	–	–	–	–	55	0.75*	68	0.47*	–	–	–	–
Texas Assessment of Academic Skills															
TAAS	2001	S 01	SS	–	–	–	–	–	–	78	0.52*	–	–	–	–
TAAS	2001	S 01	TLI	–	–	–	–	–	–	–	–	–	–	82	0.42*
Virginia Standards of Learning															
Virginia		S 00	SS	–	–	–	–	–	–	–	–	24	0.73*	–	–
Washington Assessment of Student Learning															
Wash		S 00	SS	–	–	–	–	–	–	–	–	–	–	90	0.54*
Wide Range Achievement Test															
WRAT III		F 01	NCE	–	–	–	–	–	–	44	0.32*	44	0.66*	–	–

Summary

Grade(s)	All	1	2	3	4	5	6
Number of students	4,996	69	262	804	1,102	1,565	1,194
Number of coefficients	98	2	6	17	23	29	21
Average validity	–	0.7	0.65	0.6	0.59	0.62	0.65
Overall average	0.62						

a. n = Sample size.
* Denote correlation coefficients that are statistically significant at the 0.05 level.

Table 51: External Validity Data—Star Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 7–12[a]

Test	Version	Date	Score	7		8		9		10		11		12	
				n	r	n	r	n	r	n	r	n	r	n	r
American College Testing Program															
ACT		F 01	NCE	–	–	–	–	–	–	–	–	–	–	26	0.87*
California Achievement Tests															
CAT	5th Ed.	F 01	NCE	–	–	–	–	64	0.73*	–	–	–	–	–	–
CAT	5th Ed.	F 01	SS	170	0.54*	–	–	–	–	–	–	–	–	–	–
Comprehensive Test of Basic Skills															
CTBS	4th Ed.	S 00	SS	67	0.67*	75	0.73*	–	–	–	–	–	–	–	–
CTBS	A-13	S 00	SS	–	–	31	0.65*	–	–	–	–	–	–	–	–
CTBS	A-13	S 01	SS	23	0.82*	–	–	–	–	48	0.63*	–	–	–	–
Delaware Student Testing Program															
DSTP		S 01	SS	–	–	–	–	94	0.27*	–	–	–	–	–	–
Differential Aptitude Tests															
DAT	Level 1	F 01	NCE	–	–	–	–	41	0.70*	–	–	–	–	–	–
Explore Tests															
Explore		F 01	NCE	–	–	64	0.54*	–	–	–	–	–	–	–	–
Georgia High School Graduation Test															
Georgia		S 01	NCE	–	–	–	–	–	–	–	–	–	–	23	0.71*
Indiana Statewide Testing for Educational Progress															
ISTEP		F01	NCE	–	–	–	–	51	0.57*	22	0.58*	–	–	–	–
Iowa Tests of Basic Skills															
ITBS	Form A	F 01	SS	66	0.71*	–	–	–	–	–	–	–	–	–	–
ITBS	Form K	S 01	NCE	73	0.80*	18	0.52*	–	–	–	–	–	–	–	–
ITBS	Form K	F 01	NCE	6	0.72	14	0.69*	–	–	–	–	–	–	–	–
ITBS	Form L	S 01	NCE	36	0.74*	32	0.53*	–	–	19	0.67*	32	0.84*	–	–
ITBS	Form M	S 99	NCE	–	–	5	0.89*	–	–	–	–	11	0.80*	–	–
ITBS	Form M	S 00	NCE	–	–	–	–	–	–	9	0.94*	–	–	–	–
ITBS	Form M	S 01	NCE	49	0.52*	48	0.51*	–	–	–	–	–	–	–	–
Kentucky Core Content Test															
KCCT		S 01	NCE	–	–	–	–	45	0.43*	–	–	–	–	–	–
Maryland High School Placement Test															
Maryland		S 01	NCE	–	–	–	–	47	0.60*	–	–	–	–	–	–

Table 51: External Validity Data—Star Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 7–12[a]

Test	Version	Date	Score	7		8		9		10		11		12	
				n	r	n	r	n	r	n	r	n	r	n	r
McGraw Hill Mississippi/Criterion Referenced															
McGraw		S 01	SS	–	–	–	–	73	0.56*	–	–	–	–	–	–
Metropolitan Achievement Test															
MAT	7th Ed.	F 01	NCE	5	0.8	11	0.82*	–	–	–	–	–	–	–	–
North Carolina End of Grade Tests															
NCEOG		S 01	SS	–	–	177	0.59*	–	–	–	–	–	–	–	–
Oklahoma School Testing Program Core Curriculum Tests															
Oklahoma		S 01	SS	–	–	–	–	26	0.67*	–	–	–	–	–	–
Oregon State Assessment															
Oregon		S 01	NCE	–	–	45	0.53*	–	–	–	–	–	–	–	–
PLAN															
PLAN		F 99	SS	–	–	–	–	–	–	–	–	–	–	–	0.42
PLAN		F 00	SS	–	–	–	–	–	–	–	–	40	0.28	–	–
PLAN		F 01	NCE	–	–	–	–	–	–	63	0.61*	–	–	–	–
Preliminary SAT/National Merit Scholarship Qualifying Test															
PSAT/NMSQT	NMSQT	F 00	NCE	–	–	–	–	–	–	–	–	–	–	–	0.63*
PSAT/NMSQT	NMSQT	F 01	NCE	–	–	–	–	–	–	–	–	72	0.64*	–	–
Stanford Achievement Test															
SAT9		S 98	NCE	11	0.84*	–	–	–	–	–	–	–	–	–	–
SAT9		S 99	NCE	14	0.71*	–	–	–	–	–	–	–	–	–	–
SAT9		F 00	SS	–	–	45	0.85*	–	–	–	–	–	–	–	–
SAT9		S 01	NCE	45	0.71*	105	0.81*	11	0.69*	–	–	–	–	–	–
SAT9		S 01	SS	54	0.76*	109	0.69*	19	0.27	77	0.59*	67	0.76*	71	0.65*
SAT9		F 01	SS	104	0.84*	–	–	–	–	–	–	–	–	–	–
TerraNova															
TerraNova		S 99	NCE	35	0.61*	47	0.62*	–	–	–	–	–	–	–	–
TerraNova		S 00	SS	18	0.73*	–	–	–	–	–	–	–	–	–	–
TerraNova		S 01	NCE	17	0.29	17	0.52*	–	–	–	–	–	–	–	–
TerraNova		S 01	SS	–	–	99	0.74*	–	–	–	–	–	–	–	–
TerraNova		F 01	SS	–	–	38	0.74*	–	–	–	–	–	–	–	–
Test of Achievement Proficiency															
TAP		F 01	NCE	–	–	–	–	8	0.7	7	0.7	–	–	–	–

Table 51: External Validity Data—Star Math 2.0 Correlation Coefficients (r) with External Tests Administered Prior to Spring 2002, Grades 7–12[a]

Test	Version	Date	Score	7		8		9		10		11		12	
				n	r	n	r	n	r	n	r	n	r	n	r
Texas Assessment of Academic Skills, 2001															
TAAS	2001	S 01	SS	66	0.44*	69	0.33*	–	–	–	–	–	–	–	–
Virginia Standards of Learning															
Virginia		S 00	SS	25	0.71*	–	–	–	–	–	–	–	–	–	–

Summary

Grade(s)	All	7	8	9	10	11	12
Number of students	3,066	930	1,049	479	245	222	141
Number of coefficients	66	20	19	11	7	5	4
Average validity	–	0.67	0.65	0.56	0.67	0.66	0.6
Overall average	0.64						

a. n = Sample size.
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 52: Concurrent Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Arkansas Augmented Benchmark Examination (AABE)														
AABE	S 08	SS	–	–	–	–	725	0.68*	686	0.70*	634	0.70*	297	0.66*
ACT Aspire														
ACT Aspire – Mathematics	S 14–16	SS	–	–	–	–	5212	0.78*	5005	0.76*	4796	0.78*	4311	0.77*
California Achievement Test (CAT) 5th Edition														
CAT	S 02	NCE	–	–	–	–	17	0.50*	–	–	–	–	–	–
CAT/5	F 10–11	SS	105	0.74*	166	0.64*	209	0.65*	242	0.54*	202	0.71*	186	0.66*
Canadian Achievement Test														
CAT/2	F 10–11	SS	–	–	–	–	–	–	24	0.74*	21	0.63*	–	–
Comprehensive Test of Basic Skills (CTBS)														
CTBS–A13	S 02	SS	–	–	–	–	–	–	–	–	21	0.66*	–	–
CTBS	S 02	NCE	–	–	–	–	–	–	–	–	–	–	32	0.65*

Table 52: Concurrent Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Delaware Student Testing Program (DSTP)														
DSTP	S 03	SS	–	–	–	–	258	0.72*	–	–	296	0.73*	–	–
DSTP	S 05	SS	–	–	–	–	66	0.67*	–	–	–	–	–	–
DSTP	S 06	SS	–	–	140	0.66*	58	0.85*	40	0.63*	151	0.75*	44	0.77*
Florida Comprehensive Assessment Test (FCAT)														
FCAT	S 06	SS	–	–	–	–	58	0.85*	40	0.63*	–	–	–	–
FCAT	S 06–08	SS	–	–	–	–	2,338	0.74*	2,211	0.74*	2,078	0.74*	279	0.65*
Florida Standards Assessments (FSA)														
FSA	S 15	SS	–	–	–	–	1,508	0.78*	1,944	0.79*	2,637	0.82*	1,434	0.84*
Georgia Milestones														
Milestones – Mathematics	S 15	SS	–	–	–	–	11262	0.79*	10434	0.79*	10925	0.79	6732	0.79*
Idaho Standards Achievement Test (ISAT)														
ISAT	F 02	SS	–	–	–	–	192	0.68*	188	0.75*	194	0.75*	221	0.74*
ISAT	S 03	SS	–	–	–	–	224	0.74*	209	0.83*	222	0.78*	231	0.82*
ISAT	S 07–09	SS	–	–	–	–	798	0.70*	699	0.60*	727	0.62*	217	0.69*
Iowa Test of Basic Skills (ITBS)														
ITBS–A	S 02	NCE	–	–	–	–	–	–	50	0.66*	79	0.72*	–	–
ITBS–K	S 02	SS	–	–	–	–	–	–	–	–	–	–	70	0.69*
ITBS–L	S 02	NCE	–	–	7	0.78*	23	0.57*	17	0.70*	21	0.66*	–	–
ITBS–M	S 02	NCE	14	0.56*	11	0.58*	–	–	–	–	–	–	–	–
ITBS–M	S 02	SS	–	–	–	–	17	0.72*	–	–	–	–	–	–
Kansas State Assessment Program (KSAP)														
KSAP	S 06–08	SS	–	–	–	–	915	0.59*	947	0.67*	752	0.66*	402	0.67*
Kentucky Core Content Test (KCCT)														
KCCT	S 08–10	SS	–	–	–	–	3,777	0.69*	3,115	0.70*	2,228	0.66*	1,785	0.66*
Key Stage 2 Standardised Attainment Tests (KS2 SATs)														
Maths	S 16	SS	–	–	–	–	–	–	–	–	815	0.84*	–	–
Maths	S 16	Raw	–	–	–	–	–	–	–	–	815	0.83*	–	–
McGraw Hill Mississippi/Criterion Referenced														
	S 02	SS	–	–	–	–	–	–	–	–	44	0.73*	–	–

Table 52: Concurrent Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Metropolitan Achievement Test (MAT)														
MAT– 6th Ed.	S 02	NCE	69	0.55*	–	–	–	–	–	–	–	–	–	–
MAT– 8th Ed.	S 02	SS	–	–	–	–	–	–	38	0.83*	–	–	–	–
Michigan Educational Assessment Program (MEAP) – Mathematics														
MEAP	F 04	SS	–	–	–	–	–	–	154	0.81*	–	–	–	–
MEAP	F 05	SS	–	–	–	–	71	0.75*	69	0.78*	77	0.83*	89	0.77*
MEAP	F 06	SS	–	–	–	–	162	0.72*	–	–	53	0.67*	123	0.69*
Minnesota Comprehensive Assessment (MCA)														
MCA	S 03	SS	–	–	–	–	85	0.71*	–	–	81	0.76*	–	–
MCA	S 04	SS	–	–	–	–	91	0.74*	–	–	83	0.73*	–	–
Mississippi Academic Assessment Program (MAAP)														
MAAP – Mathematics	S 16	SS	–	–	–	–	2058	0.78*	1633	0.79*	2045	0.72*	2145	0.74*
Mississippi Curriculum Test (MCT2)														
CTB	S 02	SS	–	–	–	–	–	–	10	0.62*	–	–	–	–
CTB	S 03	SS	–	–	–	–	117	0.71*	154	0.77*	119	0.78*	52	0.43*
MCT	S 03	SS	–	–	–	–	117	0.71*	154	0.77*	110	0.78*	52	0.43*
MCT2	S 08	SS	–	–	–	–	1,786	0.72*	1,757	0.72*	1,531	0.73*	1,180	0.78*
Missouri Assessment Program (MAP) Grade-Level Tests														
MAP – Mathematics	S 16	SS	–	–	–	–	4403	0.84*	4276	0.83*	4239	0.83*	2266	0.84*
New Jersey Assessment of Skills and Knowledge (NJASK)														
NJASK	S 13	SS	–	–	–	–	1,589	0.82*	1,715	0.82*	1,485	0.85*	389	0.76*
New York State Assessment Program														
NYSTP	S 13	SS	–	–	–	–	122	0.73*	–	–	–	–	–	–
North Carolina End-of-Grade (NCEOG) Test														
NCEOG	S 02	NCE	–	–	–	–	70	0.60*						
NCEOG	S 02	SS					62	0.73*						
NCEOG	S 06–08	SS	–	–	–	–	1,100	0.72*	751	0.72*	482	0.65*	202	0.77*
NCEOG	S 14	SS	–	–	–	–	9,235	0.76*	8,324	0.76*	7,866	0.77*	4,618	0.78*

Table 52: Concurrent Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
NWEA, NALT, & MAP														
	F 02	SS	–	–	–	–	81	0.75*	–	–	77	0.86*	–	–
	S 03	SS	–	–	–	–	85	0.82*	–	–	80	0.85*	–	–
	F 03	SS	–	–	77	0.69*	92	0.73*	75	0.82*	79	0.86*	–	–
	S 04	SS	–	–	80	0.72*	92	0.84*	65	0.84*	82	0.86*	–	–
	F 04	SS	–	–	–	–	63	0.53*	77	0.78*	86	0.84*	–	–
	S 05	SS	–	–	–	–	63	0.74*	80	0.87*	96	0.87*	–	–
Ohio Achievement Assessment														
OAA	S 13	SS	–	–	–	–	1,725	0.76*	1,594	0.75*	1,605	0.76*	1,601	0.69*
Ohio State Tests (OST)														
OST – Mathematics	S 16	SS	–	–	–	–	4397	0.82*	3870	0.83*	3514	0.80*	3752	0.77*
Oklahoma Core Curriculum Test (OCCT)														
OCCT	S 06	SS	–	–	–	–	77	0.71*	92	0.61*	66	0.68*	60	0.63*
Partnership for Assessment of Readiness for College and Careers (PARCC)														
PARCC	S 15	SS	–	–	–	–	4,103	0.8	4,787	0.83*	4,266	0.79*	5,050	0.8*
Pennsylvania System of School Assessment (PSSA)														
PSSA	S 02	SS	–	–	–	–	–	–	–	–	–	–	62	0.76*
PSSA	S 13	SS	–	–	–	–	87	0.76*	76	0.86*	70	0.64*	–	–
South Dakota State Test of Educational Progress (DSTEP)														
DSTEP	S 08–10	SS	–	–	–	–	2,092	0.74*	1,555	0.74*	1,309	0.72*	837	0.74*
Stanford Achievement Test (SAT9)														
SAT9	S 02	NCE	–	–	113	0.56*	39	0.83*	46	0.54*	103	0.70*	49	0.65*
SAT9	S 02	SS	20	0.76*	16	0.68*	18	0.59*	19	0.57*	71	0.49*	84	0.62*
Smarter Balanced Assessment (SBA)														
SBA	S 15	SS	–	–	–	–	608	0.85*	640	0.87*	513	0.85*	561	0.86*
SBA	S 15	SS	–	–	–	–	10,800	0.84*	10,582	0.86*	9,750	0.86*	7,852	0.86*
State of Texas Assessments of Academic Readiness Standards Test 2														
STAAR	S 12–13	SS	–	–	–	–	5,794	0.73*	6,141	0.75*	5,538	0.71*	4,437	0.75*
STAAR	S 11–14	SS	–	–	–	–	6,424	0.77*	6,138	0.76*	1,833	0.78*	5,331	0.73*

Table 52: Concurrent Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Tennessee Comprehensive Assessment Program (TCAP)														
TCAP	S 11	SS	–	–	–	–	35	0.78*	–	–	–	–	–	–
TCAP	S 12	SS	–	–	–	–	72	0.76*	98	0.69*	74	0.85*	–	–
TCAP	S 13	SS	–	–	–	–	172	0.74*	232	0.63*	286	0.68*	–	–
TerraNova														
TerraNova	S 02	NCE	7	0.66*	14	0.46*	125	0.68*	18	0.67*	17	0.79*	15	0.64*
TerraNova	F 03	SS	–	–	177	0.55*	172	0.45*	119	0.67*	160	0.78*	–	–
TerraNova	S 04	SS	–	–	150	0.75*	205	0.71*	149	0.71*	182	0.78*	–	–
Texas Assessment of Academic Achievement (TAAS)														
TAAS	S 01	SS	–	–	–	–	1,036	0.56*	1,047	0.50*	1,006	0.65*	991	0.61*
TAAS	S 02	SS	–	–	–	–	674	0.65*	669	0.63*	677	0.64*	885	0.64*
Texas Assessment of Knowledge and Skills (TAKS)														
TAKS	S 03	SS	–	–	–	–	1,134	0.63*	1,129	0.62*	1,086	0.70*	–	–
Transitional Colorado Assessment Program (TCAP)														
TCAP	S 12–13	SS	–	–	–	–	3,185	0.84*	3,211	0.88*	3,183	0.89*	3,111	0.90*
West Virginia Educational Standards Test 2														
WESTEST 2	S 12	SS	–	–	–	–	2,386	0.74*	2,725	0.75*	2,324	0.75*	1,153	0.73*
Wisconsin Forward Exam														
WI Forward – Mathematics	S 16	SS	–	–	–	–	8720	0.79*	8255	0.76*	8047	0.73*	6941	0.82*
Wisconsin Knowledge and Concepts Examination (WKCE)														
WKCE	F 06–10	SS	–	–	–	–	1,322	0.71*	1,393	0.72*	1,801	0.73*	1,175	0.75*

Summary

Grade(s)	All	1	2	3	4	5	6
Number of students	370,651	215	951	104,603	99,768	93,810	71,304
Number of coefficients	241	5	11	64	56	62	43
Average validity	–	0.65	0.64	0.72	0.73	0.75	0.72
Overall average	0.73						

a. n = Sample size.
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 53: Concurrent Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Arkansas Augmented Benchmark Examination (AABE)														
AABE	S 08	SS	99	0.56*	74	0.77*	–	–	–	–	–	–	–	–
ACT														
ACT – Mathematics	S 08– 15	SS	–	–	–	–	14	0.54*	177	0.47*	1278	0.66*	26	–0.04
ACT Aspire														
ACT Aspire – Mathematics	S 14–16	SS	3351	0.81*	3377	0.82*	5083	0.65*	3981	0.76*	–	–	–	–
California Achievement Test (CAT) 5th Edition														
CAT/5	F 10–11	SS	166	0.73*	129	0.64*	52	0.71*	33	0.68*	–	–	–	–
Delaware Student Testing Program (DSTP)														
DSTP	S 03	SS	–	–	254	0.78*	–	–	–	–	–	–	–	–
Florida Comprehensive Assessment Test (FCAT)														
FCAT	S 02	SS	–	–	–	–	–	–	51	0.64*	57	0.66*	38	0.75*
FCAT	S 06–08	SS	195	0.65*	89	0.60*	–	–	–	–	–	–	–	–
Florida Standards Assessments (FSA)														
FSA	S 15	SS	1,211	0.82*	936	0.71*	–	–	–	–	–	–	–	–
Georgia Milestones														
Milestones – Mathematics	S 15	SS	5877	0.77*	6049	0.74*	–	–	–	–	–	–	–	–
Idaho Standards Achievement Test (ISAT)														
ISAT	F 02	SS	206	0.81*	170	0.81*	–	–	–	–	–	–	–	–
ISAT	S 03	SS	227	0.85*	174	0.82*	–	–	–	–	–	–	–	–
ISAT	S 06–08	SS	289	0.71*	328	0.77*	–	–	–	–	–	–	–	–
Iowa Test of Basic Skills (ITBS)														
ITBS–M	S 02	SS	37	0.40*	–	–	–	–	–	–	–	–	–	–
Kansas State Assessment Program (KSAP)														
KSAP	S 06–08	SS	271	0.74*	137	0.75*	–	–	–	–	–	–	–	–
Kentucky Core Content Test (KCCT)														
KCCT	S 08–10	SS	788	0.68*	362	0.64*	–	–	–	–	–	–	–	–
Measures of Academic Progress (MAP)														
MAP	S 15	SS	413	0.82	646	0.82	–	–	–	–	–	–	–	–

Table 53: Concurrent Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Michigan Educational Assessment Program (MEAP) – Mathematics														
MEAP	F 05	SS	65	0.72*	71	0.80*	–	–	–	–	–	–	–	–
MEAP	F 06	SS	122	0.84*	123	0.58*	–	–	–	–	–	–	–	–
Mississippi Academic Assessment Program (MAAP)														
MAAP – Mathematics	S 16	SS	1417	0.73*	1185	0.70*	–	–	–	–	–	–	–	–
Mississippi Curriculum Test (MCT2)														
MCT2	S 08	SS	721	0.66*	549	0.71*	–	–	–	–	–	–	–	–
Missouri Assessment Program (MAP) Grade-Level Tests														
MAP – Mathematics	S 16	SS	1874	0.76*	1294	0.73*	–	–	–	–	–	–	–	–
New Standards Reference Mathematics Exam (Rhode Island)														
NRSME	S 02	SS	–	–	–	–	–	–	–	–	67	0.67*	9	0.66*
North Carolina End-of-Grade (NCEOG) Test														
NCEOG	S 06–08	SS	216	0.70*	39	0.81*	–	–	–	–	–	–	–	–
NCEOG	S 14	SS	3,947	0.73*	3,302	0.72*	–	–	–	–	–	–	–	–
New Jersey Assessment of Skills and Knowledge (NJASK)														
NJASK	S 13	SS	620	0.79*	611	0.78*	–	–	–	–	–	–	–	–
Ohio Achievement Assessment														
OAA	S 13	SS	1,412	0.65*	1,380	0.65*	–	–	–	–	–	–	–	–
Ohio State Tests (OST)														
OST – Mathematics	S 16	SS	3412	0.77*	2883	0.73*	–	–	–	–	–	–	–	–
Ohio Proficiency Test (OPT)														
OPT	S 02	SS	–	–	–	–	23	0.67*	26	0.40*	24	0.77*	24	0.69*
Oklahoma Core Curriculum Test (OCCT)														
OCCT	S 06	SS	55	0.63*	68	0.70*	–	–	–	–	–	–	–	–
Otis Lennon School Ability Test (OLSAT)														
OLSAT	S 02	NCE	–	–	–	–	–	–	12	0.36	13	0.91*	6	0.72*
Palmetto Achievement Challenge Test (PACT), 2001														
PACT	S 02	SS	–	–	161	0.72*	–	–	–	–	–	–	–	–
Partnership for Assessment of Readiness for College and Careers (PARCC)														
PARCC	S 15	SS	4,368	0.77*	4,196	0.75*	–	–	–	–	–	–	–	–

Table 53: Concurrent Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
South Dakota State Test of Educational Progress (DSTEP)														
DSTEP	S 08–10	SS	525	0.73*	535	0.73*	–	–	–	–	–	–	–	–
Smarter Balanced Assessment (SBA)														
SBA	S 15	SS	569	0.82*	432	0.79*	–	–	–	–	55	0.52	–	–
SBA	S 15	SS	6,344	0.86*	5,424	0.83*	–	–	–	–	–	–	–	–
State of Texas Assessments of Academic Readiness Standards Test 2														
STAAR	S 12–13	SS	4,171	0.71*	3,379	0.68*	–	–	–	–	–	–	–	–
STAAR	S 11–14	SS	4,437	0.74*	–	–	–	–	–	–	–	–	–	–
Texas Assessment of Academic Achievement (TAAS)														
TAAS	S 01	SS	892	0.60*	825	0.67*	–	–	–	–	–	–	–	–
TAAS	S 02	SS	768	0.62*	809	0.68*	–	–	–	–	–	–	–	–
Texas Assessment of Academic Skills (TAAS), 2001														
TAAS	S 02	TLI	–	–	–	–	163	0.69*	–	–	–	–	–	–
Transitional Colorado Assessment Program (TCAP)														
TCAP	S 12–13	SS	3,173	0.90*	3,114	0.88*	–	–	–	–	–	–	–	–
West Virginia Educational Standards Test 2														
WESTEST 2	S 12	SS	1,184	0.76*	1,215	0.69*	–	–	–	–	–	–	–	–
Wisconsin Forward Exam														
WI Forward – Mathematics	S 16	SS	6855	0.74	6355	0.7	–	–	–	–	–	–	–	–
Wisconsin Knowledge and Concepts Examination (WKCE)														
WKCE	F 06–10	SS	640	0.79*	767	0.76*	–	–	248	0.73*	–	–	–	–

Summary

Grade(s)	All	7	8	9	10	11	12
Number of students	123,819	60,917	51,442	5,335	4,528	1,494	103
Number of coefficients	95	36	36	5	7	6	5
Average validity	–	0.73	0.74	0.65	0.58	0.7	0.56
Overall average	0.71						

a. n = Sample size.
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 54: Predictive Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Arkansas Augmented Benchmark Examination (AABE)														
AABE	F 07	SS	–	–	–	–	1,196	0.69*	1,128	0.67*	994	0.73*	638	0.71*
ACT Aspire														
ACT Aspire	S 14	SS	–	–	–	–	373	0.77*	392	0.67*	380	0.61*	359	0.70*
ACT Aspire – Mathematics	F 13–S 16	SS	–	–	–	–	5117	0.80*	4994	0.78*	5096	0.78*	4090	0.78*
Delaware Student Testing Program (DSTP)														
DSTP	F 02	SS	–	–	–	–	191	0.70*	–	–	228	0.70*	–	–
DSTP	F 04	SS	–	–	–	–	171	0.67*	–	–	–	–	–	–
DSTP	W 05	SS	–	–	–	–	149	0.76*	–	–	–	–	–	–
DSTP	S 05	SS	–	–	–	–	132	0.64*	172	0.63*	185	0.62*	–	–
DSTP	F 05	SS	–	–	206	0.64*	219	0.66*	249	0.67*	265	0.68*	–	–
DSTP	W 05	SS	–	–	242	0.61*	226	0.61*	269	0.62*	277	0.68*		
Florida Comprehensive Assessment Test (FCAT)														
FCAT	F 05	SS	–	–	–	–	54	0.79*	42	0.69*	–	–	–	–
FCAT	F 05–07	SS	–	–	–	–	5,292	0.74*	5,020	0.73*	4,895	0.77*	1,015	0.66*
Florida Standards Assessments (FSA)														
FSA	S 15	SS	–	–	–	–	4,188	0.81*	4,133	0.82*	4,107	0.81*	1,398	0.84*
Georgia Milestones														
Milestones – Mathematics	F 14–S 15	SS	–	–	–	–	8279	0.82*	7868	0.81*	7802	0.82*	6965	0.80*
Idaho Standards Achievement Test (ISAT)														
ISAT	F 08–10	SS	–	–	–	–	1,875	0.67*	1,908	0.63*	2,312	0.69*	1,809	0.73*
Iowa Assessment														
IA	F 12	SS	–	–	–	–	770	0.67*	885	0.65*	896	0.56*	732	0.48*
IA	W 12	SS	–	–	–	–	1,299	0.61*	997	0.62*	923	0.58*	918	0.64*
IA	S 12	SS	–	–	–	–	299	0.66*	301	0.67*	268	0.62*	204	0.62*
Kentucky Core Content Test (KCCT)														
KCCT	F 07–09	SS	–	–	–	–	5,821	0.68*	5,325	0.67*	4,199	0.66*	3,172	0.63*
Kentucky Performance Rating for Educational Progress (K-PREP)														
K-PREP	S 12	SS	–	–	–	–	557	0.82*	556	0.87*	537	0.85*	43	0.66*
Louisiana Educational Assessment Program (LEAP 2025)														
LEAP 2025 – Mathematics	F 15–S 16	SS	–	–	–	–	1965	0.80*	1964	0.80*	1653	0.77*	703	0.80*

Table 54: Predictive Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 1–6[a]

			1		2		3		4		5		6	
Test Form	**Date**	**Score**	**n**	**r**	**n**	**r**	**n**	**r**	**n**	**r**	**n**	**r**	**n**	**r**
Maine Educational Assessment (MEA)														
MEA – Mathematics	F 15–S 16	SS	–	–	–	–	139	0.81*	142	0.77*	157	0.72*	158	0.74*
Michigan Educational Assessment Program (MEAP)														
MEAP	F 04	SS	–	–	–	–	–	–	64	0.70*	74	0.85*	81	0.74*
MEAP	W 05	SS	–	–	–	–	–	–	65	0.80*	75	0.87*	42	0.72*
MEAP	S 05	SS	–	–	–	–	66	0.63*	65	0.73*	76	0.83*	84	0.71*
Michigan Student Test of Educational Progress (M-STEP)														
M-STEP	S 15	SS	–	–	–	–	783	0.85*	758	0.85*	345	0.84*	644	0.84*
Georgia Milestones – English Language Arts														
Milestones	S 15	SS	–	–	–	–	814	0.86*	721	0.84*	845	0.83*	471	0.8*
Minnesota Comprehensive Assessment (MCA)														
MCA	F 02	SS	–	–	–	–	81	0.64*	–	–	78	0.72*	–	–
MCA	W 03	SS	–	–	–	–	86	0.66*	–	–	81	0.77*	–	–
MCA	F 03	SS	–	–	–	–	87	0.53*	–	–	79	0.69*	–	–
MCA	W 04	SS	–	–	–	–	93	0.60*	–	–	82	0.75*	–	–
Mississippi Academic Assessment Program (MAAP)														
MAAP – Mathematics	F 15–S 16	SS	–	–	–	–	2390	0.79*	1937	0.70*	1686	0.69*	1662	0.78*
Mississippi Curriculum Test (MCT2)														
MCT	F 02	SS	–	–	–	–	48	0.64*	33	0.82*	73	0.80*	–	–
MCT	F 03	SS	–	–	–	–	109	0.51*	164	0.72*	156	0.69*	–	–
MCT2	F 07	SS	–	–	–	–	2,989	0.69*	3,022	0.70*	2,796	0.72*	2,741	0.74*
Missouri Assessment Program (MAP) Grade-Level Tests														
MAP – Mathematics	F 15–S 16	SS	–	–	–	–	3846	0.86*	3836	0.84*	3872	0.84*	2930	0.84*
New York State Assessment Program														
NYSTP	F 12	SS	–	–	–	–	290	0.60*	–	–	–	–	–	–
North Carolina End-of-Grade (NCEOG) Test														
NCEOG	F 05–07	SS	–	–	–	–	2,494	0.73*	2,008	0.70*	1,096	0.69*	830	0.70*
NCEOG	S 14	SS	–	–	–	–	29,878	0.71*	28,659	0.73*	27,366	0.73*	15,420	0.74*
NWEA NALT & MAP														
	F 02	–	–	–	–	–	80	0.65*	–	–	77	0.86*	–	–
	W 03	–	–	–	–	–	85	0.78*	–	–	80	0.90*	–	–
	F 03	–	–	–	–	–	86	0.68*	69	0.81*	78	0.87*	–	–
	W 04	–	–	–	–	–	92	0.80*	68	0.80*	81	0.93*	–	–

Table 54: Predictive Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Oklahoma Core Curriculum Test (OCCT)														
OCCT	F 05	SS	–	–	–	–	87	0.71*	88	0.61*	77	0.55*	83	0.56*
Ohio Achievement Assessment														
OAA	F 12	SS	–	–	–	–	47	0.82*	43	0.76*	34	0.71*	32	0.61*
Ohio State Tests (OST)														
OST – Mathematics	F 15–S 16	SS	–	–	–	–	3846	0.83*	3588	0.84*	3255	0.81*	3371	0.80*
Partnership for Assessment of Readiness for College and Careers (PARCC)														
PARCC	S 15	SS	–	–	–	–	3,635	0.83*	4,008	0.83*	3,653	0.8*	4,150	0.82*
Pennsylvania System of School Assessment (PSSA)														
PSSA	S 12	SS	–	–	–	–	92	0.82*	84	0.88*	74	0.7*	–	–
PSSA	F 12	SS	–	–	–	–	87	0.79*	74	0.81*	72	0.59*	–	–
PSSA	F 12	SS	–	–	–	–	84	0.82*	70	0.79*	73	0.65*	–	–
PSSA	W 13	SS	–	–	–	–	86	0.78*	74	0.81*	72	0.66*	–	–
PSSA	W 13	SS	–	–	–	–	86	0.8*	75	0.85*	75	0.61*	–	–
PSSA	S 13	SS	–	–	–	–	85	0.76*	74	0.84*	73	0.65*	–	–
PSSA	S 13	SS	–	–	–	–	85	0.78*	69	0.84*	71	0.71*	–	–
PSSA	S 15	SS	–	–	–	–	580	0.85*	717	0.84*	606	0.82*	575	0.85*
South Carolina College-and Career-Ready Assessments (SC READY)														
SC READY – Mathematics	F 15–S 16	SS	–	–	–	–	2224	0.82*	2047	0.79*	1428	0.82*	1092	0.79*
South Dakota State Test of Educational Progress (DSTEP)														
DSTEP	F 07–09	SS	–	–	–	–	3,886	0.73*	3,665	0.75*	3,084	0.72*	2,328	0.75*
Smarter Balanced Assessment (SBA)														
SBA	F 14	SS	–	–	–	–	608	0.82*	640	0.81*	513	0.83*	561	0.82*
SBA	W 14	SS	–	–	–	–	608	0.83*	640	0.84*	513	0.83*	561	0.84*
SBA	S 15	SS	–	–	–	–	8,593	0.87*	8,571	0.88*	8,595	0.88*	8,575	0.88*
STAR Math														
STAR–M	F 01	SS	–	–	–	–	1,036	0.61*	1,047	0.63*	1,006	0.65*	991	0.65*
STAR–M	F 05	SS	2,605	0.50*	7,195	0.63*	11,716	0.67*	13,295	0.69*	10,343	0.70*	6,823	0.75*
STAR–M	F 06	SS	4,687	0.58*	12,464	0.62*	16,474	0.66*	17,161	0.70*	16,181	0.71*	12,026	0.73*
STAR–M	F 05	SS	1,147	0.51*	3,181	0.62*	4,894	0.67*	5,254	0.70*	2,164	0.69*	1,474	0.74*
STAR–M	F 05	SS	1,147	0.42*	3,181	0.57*	4,894	0.62*	5,254	0.64*	2,164	0.73*	1,474	0.80*
STAR–M	S 06	SS	1,147	0.66*	3,181	0.69*	4,894	0.73*	5,254	0.74*	2,164	0.73*	1,474	0.80*
STAR–M	S 06	SS	1,147	0.62*	3,181	0.63*	4,894	0.69*	5,254	0.70*	2,164	0.71*	1,474	0.78*

Table 54: Predictive Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
State of Texas Assessments of Academic Readiness Standards Test 2														
STAAR	F 11–12	SS	–	–	–	–	4,788	0.75*	4,945	0.76*	4,740	0.76*	4,353	0.74*
STAAR	S 14–15	SS	–	–	–	–	4,744	0.8*	4,613	0.77*	3,878	0.77*	4,878	0.74*
Tennessee Comprehensive Assessment Program (TCAP)														
TCAP	F 10	SS	–	–	–	–	329	0.51*	305	0.58*	307	0.63*	–	–
TCAP	F 11	SS	–	–	–	–	328	0.58*	229	0.60*	406	0.64*	–	–
TCAP	F 12	SS	–	–	–	–	591	0.62*	522	0.65*	649	0.67*	290	0.75*
TCAP	S 14	SS	–	–	–	–	127	0.82*	122	0.87*	–	–	–	–
Texas Assessment of Academic Achievement (TAAS)														
TAAS	F 01	SS	–	–	–	–	1,036	0.51*	1,047	0.42*	1,006	0.60*	991	0.61*
Texas Assessment of Knowledge and Skills (TAKS)														
TAKS	F 02	SS	–	–	–	–	262	0.64*	135	0.49*	228	0.70*	646	0.69*
TerraNova														
TerraNova	F 03	–	–	–	117	0.69*	165	0.58*	116	0.75*	154	0.54*	–	–
TerraNova	W 04	–	–	–	128	0.58*	197	0.47*	120	0.71*	173	0.77*	–	–
West Virginia Educational Standards Test 2														
WESTEST 2	F 11	SS	–	–	–	–	2,447	0.75*	2,536	0.77*	2,298	0.78*	1,533	0.77*
Wisconsin Forward Exam														
WI Forward – Mathematics	F 15– S 16	SS	–	–	–	–	895	0.81*	800	0.79*	785	0.73*	711	0.84*
Wisconsin Knowledge and Concepts Examination (WKCE)														
WKCE	S 05–09	SS	–	–	–	–	4,645	0.66*	4,980	0.68*	5,345	0.74*	4,702	0.75*

Summary

Grade(s)	All	1	2	3	4	5	6
Number of students	662,040	11,880	33,076	176,784	175,330	152,693	112,277
Number of coefficients	285	6	10	77	69	74	49
Average validity	–	0.55	0.63	0.72	0.74	0.73	0.74
Overall average	0.72						

a. n = Sample size.
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 55: Predictive Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Arkansas Augmented Benchmark Examination (AABE)														
AABE	F 07	SS	369	0.67*	296	0.76*	–	–	–	–	–	–	–	–
ACT														
ACT – Mathematics	F 07–S 15	SS	–	–	–	–	68	0.59*	1368	0.53*	4800	0.74*	92	0.43*
ACT Aspire														
ACT Aspire	S 14	SS	376	0.67*	349	0.79*	–	–	–	–	–	–	–	–
ACT Aspire – Mathematics	F 13–S 16	SS	4065	0.80*	4046	0.82*	5358	0.72*	4815	0.78*	–	–	–	–
Delaware Student Testing Program (DSTP)														
DSTP	F 02	SS	242	0.74*	–	–	–	–	–	–	–	–	–	–
DSTP	S 05	SS	227	0.71*	175	0.75*	–	–	–	–	–	–	–	–
Florida Comprehensive Assessment Test (FCAT)														
FCAT	F 05–07	SS	783	0.72*	336	0.70*	–	–	–	–	–	–	–	–
Florida Standards Assessments (FSA)														
FSA	S 15	SS	1,267	0.83*	978	0.73*	–	–	–	–	–	–	–	–
Georgia Milestones														
Milestones – Mathematics	F 14–S 15	SS	6743	0.79*	7088	0.76*	–	–	–	–	–	–	–	–
Idaho Standards Achievement Test (ISAT)														
ISAT	F 05–07	SS	588	0.75*	484	0.75*	–	–	–	–	–	–	–	–
Iowa Assessment														
IA	F 12	SS	809	0.61*	787	0.65*	–	–	–	–	–	–	–	–
IA	W 12	SS	620	0.66*	470	0.73*	–	–	–	–	–	–	–	–
IA	S 12	SS	172	0.67*	164	0.67*	–	–	–	–	–	–	–	–
Kentucky Core Content Test (KCCT)														
KCCT	F 07–09	SS	1,789	0.65*	1,153	0.59*	–	–	–	–	–	–	–	–
Kentucky Performance Rating for Educational Progress (K-PREP)														
K-PREP	S 12	SS	46	0.68*	323	0.78*	–	–	–	–	–	–	–	–
Louisiana Educational Assessment Program (LEAP 2025)														
LEAP 2025 – Mathematics	F 15–S 16	SS	865	0.82*	563	0.74*	–	–	–	–	–	–	–	–

Table 55: Predictive Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Maine Educational Assessment (MEA)														
MEA – Mathematics	F 15–S 16	SS	138	0.70*	161	0.61*	–	–	–	–	–	–	–	–
Michigan Educational Assessment Program (MEAP)														
MEAP	F 04	SS	56	0.78*	–	–	–	–	–	–	–	–	–	–
MEAP	W 05	SS	56	0.78*	–	–	–	–	–	–	–	–	–	–
MEAP	S 05	SS	37	0.86*	–	–	–	–	–	–	–	–	–	–
Michigan Student Test of Educational Progress (M-STEP)														
M-STEP	S 15	SS	1053	0.84*	677	0.8*	–	–	–	–	–	–	–	–
Georgia Milestones – English Language Arts														
Milestones	S 15	SS	453	0.8*	463	0.77*	–	–	–	–	–	–	–	–
Mississippi Academic Assessment Program (MAAP)														
MAAP – Mathematics	F 15–S 16	SS	1644	0.77*	1635	0.75*	–	–	–	–	–	–	–	–
Mississippi Curriculum Test (MCT2)														
MCT2	F 07	SS	2,127	0.71*	2,190	0.70*	–	–	–	–	–	–	–	–
Missouri Assessment Program (MAP) Grade-Level Tests														
MAP – Mathematics	F 15–S 16	SS	2734	0.74*	2224	0.73*	–	–	–	–	–	–	–	–
North Carolina End-of-Grade (NCEOG) Test														
NCEOG	F 05–07	SS	443	0.78*	397	0.71*	–	–	–	–	–	–	–	–
NCEOG	S 14	SS	1,267	0.83*	978	0.73*	–	–	–	–	–	–	–	–
Oklahoma Core Curriculum Test (OCCT)														
OCCT	F 05	SS	74	0.57*	70	0.67*	–	–	–	–	–	–	–	–
Ohio Achievement Assessment														
OAA	F 12	SS	60	0.63*	45	0.49*	–	–	–	–	–	–	–	–
Ohio State Tests (OST)														
OST – Mathematics	F 15–S 16	SS	3029	0.80*	2593	0.76*	–	–	–	–	–	–	–	–
Partnership for Assessment of Readiness for College and Careers (PARCC)														
PARCC	S 15	SS	4,066	0.8*	3,748	0.76*	–	–	–	–	–	–	–	–

Table 55: Predictive Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Pennsylvania System School Assessment (PSSA)														
PSSA – Mathematics	F 14–S 15	SS	532	0.83*	426	0.80*	–	–	–	–	–	–	–	–
South Carolina College-and Career-Ready Assessments (SC READY)														
SC READY – Mathematics	F 15–S 16	SS	1077	0.78*	1041	0.76*	–	–	–	–	–	–	–	–
South Dakota State Test of Educational Progress (DSTEP)														
DSTEP	F 07–09	SS	1,851	0.74*	1,522	0.75*	–	–	–	–	–	–	–	–
Smarter Balanced Assessment (SBA)														
SBA	F 14	SS	569	0.81*	432	0.79*	–	–	–	–	55	0.5	–	–
SBA	W 14	SS	569	0.81*	432	0.77*	–	–	–	–	55	0.59	–	–
SBA	S 15	SS	4,066	0.8*	3,748	0.76*	–	–	–	–	–	–	–	–
STAR Math														
STAR–M	F 01	–	892	0.72*	825	0.78*	–	–	–	–	–	–	–	–
STAR–M	F 05	–	3,551	0.75*	2,693	0.76*	668	0.79*	508	0.79*	572	0.79*	378	0.76*
STAR–M	F 06	–	7,564	0.76*	7,122	0.77*	1,017	0.78*	876	0.76*	693	0.83*	507	0.77*
STAR–M	F 05	–	1,191	0.75*	127	0.84*	215	0.78*	213	0.83*	164	0.75*	–	–
STAR–M	F 05	–	1,191	0.71*	127	0.77*	215	0.78*	213	0.81*	164	0.75*	–	–
STAR–M	S 06	–	1,191	0.79*	127	0.82*	215	0.80*	213	0.85*	164	0.79*	–	–
STAR–M	S 06	–	1,191	0.77*	127	0.82*	215	0.76*	213	0.82*	164	0.77*	–	–
State of Texas Assessments of Academic Readiness Standards Test 2														
STAAR	F 11–12	SS	4,177	0.72*	3,508	0.72*	–	–	–	–	–	–	–	–
STAAR	S 14–15	SS	4,350	0.76*	–	–	–	–	–	–	–	–	–	–
Tennessee Comprehensive Assessment Program (TCAP)														
TCAP	F 12	SS	273	0.80*	169	0.59*	–	–	–	–	–	–	–	–
Texas Assessment of Academic Achievement (TAAS)														
TAAS	F 01	SS	892	0.59*	825	0.67*	–	–	–	–	–	–	–	–
Texas Assessment of Knowledge and Skills (TAKS)														
TAKS	F 02	SS	564	0.74*	562	0.74*	–	–	–	–	–	–	–	–
West Virginia Educational Standards Test 2														
WESTEST 2	F 11	SS	1,437	0.78*	1,377	0.72*	–	–	–	–	–	–	–	–

Table 55: Predictive Validity Data—Star Math Correlation Coefficients (r) with External Tests Administered Between 2002 and 2016, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Wisconsin Forward Exam														
WI Forward – Mathematics	F 15–S 16	SS	667	0.74*	635	0.73*	–	–	–	–	–	–	–	–
Wisconsin Knowledge and Concepts Examination (WKCE)														
WKCE	S 05–09	SS	1,883	0.79*	1,742	0.76*	–	–	289	0.76*	–	–	–	–

Summary

Grade(s)	All	7	8	9	10	11	12
Number of students	160,323	75,876	59,960	7,971	8,708	6,831	977
Number of coefficients	126	51	46	8	9	9	3
Average validity	–	0.75	0.74	0.75	0.77	0.72	0.65
Overall average	0.74						

a. n = Sample size.
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Relationship of Star Math Scores to Teacher Ratings

In order to have a common measure of each student's math skills independent of Star Math, Renaissance Learning constructed two 12-item checklists for teachers to use during the 2014 norming study.

On this worksheet, teachers were asked to rate each student's ability to complete a wide range of tasks related to developing math skills. The intent of this checklist was to provide teachers with a single, brief instrument they could use to rate any student.

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

The Rating Instruments

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

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

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

Grade 1–5 Math Skills Rating Worksheet

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

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

Grade 6–12 Math Skills Rating Worksheet

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

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

Participating teachers were asked to complete the following rating checklist for all students in their math class:

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

Psychometric Properties of the Skills Ratings

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

The resulting Rasch scores ranged from –14.47 to 11.1. The lower value corresponds to students in grades 1 to 5 rated as possessing none of the math skills, and the higher value corresponds to students in grades 6–12 rated as possessing all of them. Table 56 lists data about the psychometric properties of the rating scale, overall and by grade, including the correlations between skills ratings and Star Math Scaled Scores. The internal consistency reliability of the rating scale was estimated as 0.93, using Cronbach's alpha.

Table 56: Psychometric Characteristics of the Skills Rating Scale and its Relationship to Scaled Scores, by Grade

Grade	N	Skills Rating		STAR Math Scaled Score		Correlation of Skills Ratings and Scaled Scores[a]
		Mean	S.D.	Mean	S.D.	
1	1,916	–6.60	2.95	385	89	0.40*
2	2,043	–3.67	2.41	503	84	0.47*
3	1,817	0.04	3.06	589	87	0.52*
4	1,820	1.26	2.83	651	90	0.58*
5	2,072	2.97	2.84	713	97	0.50*
6	1,637	5.5	2.07	763	100	0.44*
7	1,465	5.57	2.18	785	109	0.50*
8	1,639	6.96	2.5	811	117	0.54*
9	1,036	6.88	2.87	798	110	0.52*
10	688	8.78	2.38	824	119	0.38*
11	737	9.81	2.3	847	123	0.39*
12	456	10.03	2.05	876	127	0.42*
Overall	17,326	2.42	5.6	672	177	0.85*

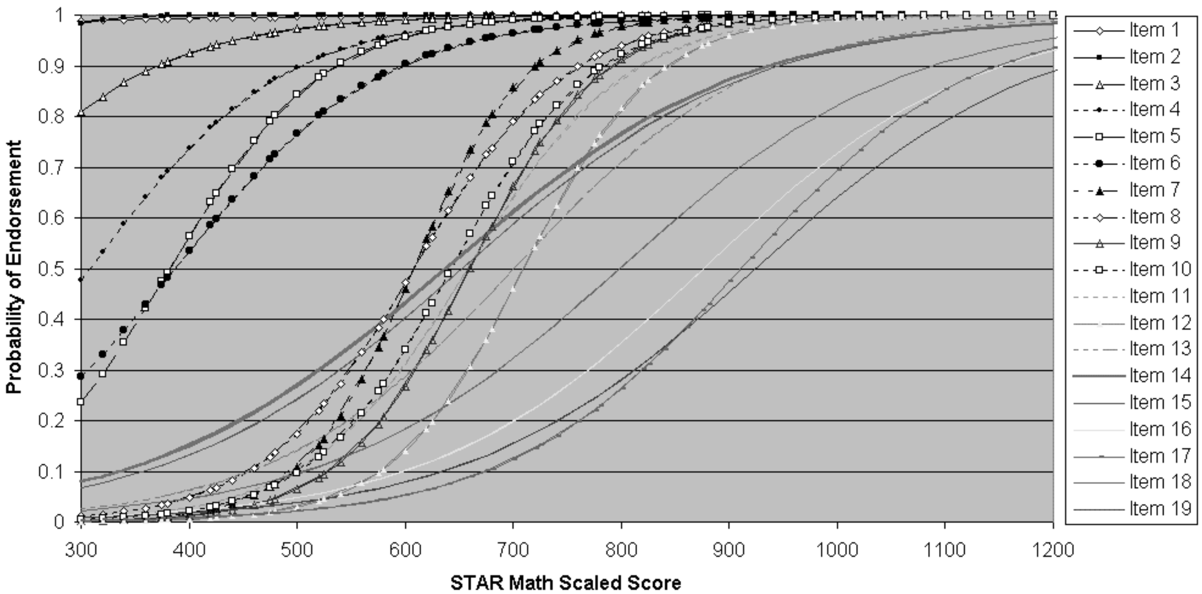
a. Asterisks denote correlation coefficients that are statistically significant at the 0.05 level.

Relationship of Star Math Scaled Scores to Math Skills Ratings

As the data in Table 56 show, the mean ratings increased directly with grade, from 6.6 at grade 1 to 10.03 at grade 12. The correlation between the skills ratings and Star Math Scaled Scores was significant at every grade level. The overall correlation was 0.85, indicating a substantial degree of relationship between the computer-adaptive Star Math test and teachers' ratings of their students' math skills.

Figure 6 displays the relationships of each of the nineteen rating scale items to Star Math Scaled Scores. These relationships were obtained by fitting mathematical models to the response data for each of the rating items. Each of the curves in the figure is a graphical depiction of the respective model. As the curves show, the proportion of students rated as possessing each of the 19 rated skills increases with the Star Math Scaled Score.

Figure 6: The Relationship of Teachers' Ratings of Student Math Skills to Star Math Scaled Scores



The relative positions of the curves provide one indication of the relative difficulty of the 19 rated skills. The rating items' Rasch difficulty parameters, displayed in Table 57, provide a somewhat different indication; the skills rating items are listed in the table from easiest to most difficult, by Rasch difficulty. The first column of Table 57 indicates the relative difficulty of the nineteen rating items, where relative difficulty 1 is the easiest and 19 is most difficult. The second and third columns list the item numbers and text of the skills rating items. The fourth column lists the Rasch difficulty scale value for each item.

The fifth column lists the correlations between students' ratings and their Star Math Scaled Scores.

Table 57: The Nineteen Rating Scale Items Listed in Order of Difficulty with Rasch Difficulty Parameters

Relative Difficulty	Item	Rating Scale Item	Rasch Difficulty	Correlation with Scaled Score[a]
Easiest	1	Identify the longest pencil among 3 pencils of different lengths.	–14.58	0.06*
↓	2	Add 2 to 4.	–14.30	0.09*
	3	State how many cents a dime is worth.	–10.28	0.26*
	4	Determine the number that shows “ones” in 162.	–7.26	0.43*
	5	Subtract 7 from 35.	–6.12	0.55*
	6	Determine the number that follows in the sequence 2, 6, 10, 14, _____.	–5.42	0.49*
	7	Divide 18 by 3.	–1.85	0.71*
	8	Write 78,318 in expanded form.	1.22	0.67*
	10	Solve the problem 4/9 + 8/9.	2.09	0.70*
	9	Read aloud the word name for 0.914.	2.51	0.70*
	11	Translate the statement “36 divided by a number is 12” into an equation.	2.59	0.67*
	12	Divide 11,540 by 577.	3.89	0.68*
	14	Solve the problem “14 is 50% of what number?”	4.54	0.40*
	15	Solve a word problem requiring the calculation of 80% of 112.	4.75	0.34*
	13	Solve a word problem requiring the calculation of proportions.	5.12	0.35*
	18	Calculate vertical and supplementary angles.	6.85	0.35*
	16	Simplify the expression (x + 1)(x + 4).	8.1	0.37*
	19	Determine 6^{-2}	9.03	0.36*
Most Difficult	17	Solve the equation $x^2 = 16x$.	9.12	0.33*

a. Asterisks denote correlation coefficients that are statistically significant at the 0.05 level.

Notice that the first two rating scale items (“Identify the longest pencil among 3 pencils of different lengths” and “Add 2 to 4”) had extremely low Rasch difficulty indices, and correlations with Scaled Scores that were near zero. As can be seen in Figure 6, these items were endorsed for nearly 100% of the students, regardless of their Star Math Scaled Scores.

As a result, they did not discriminate among students with high and low levels of developed math ability, as measured by the Star Math test.

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

Teachers' responses to items 7–12 suggest that their corresponding math tasks are considerably more difficult for their students to complete. This is reflected both in their Rasch difficulty parameters in Table 57 and in Figure 6. The figure suggests that mastery of these skills occurs between 700 and 800 on the Star Math Score Scale. The slopes of the curves for these are all steep relative to other skills items, suggesting that these skills develop rapidly, compared to the others. The correlations between these items and Scaled Scores support this hypothesis, as items 7–12 show the highest correlations with Star Math Scaled Scores.

Items 13–19 measure the most difficult of the skills. This is indicated by their Rasch difficulty parameters in Table 57 and is also confirmed by the locations at which 80% mastery occurs, illustrated in Figure 6, which suggests that these skills develop much later than all others. In fact, all students may not master these skills. Moreover, all of these items have only moderate correlations with Star Math Scaled Scores, suggesting that growth of these skills is relatively gradual.

Linking Star and State Assessments: Comparing Student- and School-Level Data

With an increasingly large emphasis on end-of-the-year summative state tests, many educators seek out informative and efficient means of gauging student performance on state standards—especially those hoping to make instructional decisions before the year-end assessment date.

For many teachers, this is an informal process in which classroom assessments are used to monitor student performance on state standards. While this may be helpful, such assessments may be technically inadequate when compared to more standardized measures of student performance. Recently the assessment scale associated with Star Math has been linked to the scales used for summative mathematics tests in nearly every state in the US. Linking Star Math assessments to state tests allows educators to reliably predict student performance on their state assessment using Star Math scores. More specifically, it places teachers in a position to identify

- which students are on track to succeed on the year-end summative state test, and
- which students might need additional assistance to reach proficiency.

Educators using Star Math assessments can access Star Performance Reports that allow access to students' Pathway to Proficiency. These reports indicate whether individual students or groups of students (by class, grade, or demographic characteristics) are likely to be on track to meet a particular state's criteria for mathematics proficiency. In other words, these reports allow instructors to evaluate student progress toward proficiency and make data-based instructional decisions well in advance of the annual state tests. Additional reports automatically generated by Star Math help educators screen for later difficulties and progress monitor students' responsiveness to interventions.

An overview of two methodologies used for linking Star Math to state assessments is provided in the following sections.

Methodology Comparison

Renaissance Learning has developed linkages between Star Math Scaled Scores and scores on the accountability tests of most states. Depending on the kind of data available for such linking, these linkages have been accomplished using one of two different methods. One method used student-level data, where both Star and state test scores were available for the same students. The other method used school-level data; this method was applied when approximately 100% of students in a school had taken Star Math, but individual students' state test scores were not available.

Student-Level Data

Using individual data to link scores between distinct assessments is commonly used when student-level data are readily available for both assessments. In this case, the distribution of standardized scores on one test (e.g. percentile ranks) may be compared to the distribution of standardized scores on another test in an effort to establish concordance. When available, individual state test data for linking purposes allowed for the comparison of Star assessments to state test scores. Star test comparison scores were obtained within an eight-week window around the median state test date (+/–4 weeks).

Typically, states classify students into one of three, four, or five performance levels on the basis of cut scores (e.g. Below Basic, Basic, Proficient, or Advanced).

After each testing period, a distribution of students falling into each of these categories will always exist (e.g. 30% in Basic, 25% in Proficient, etc.). Because Star data were available for the same students who completed the state test, the distributions could be linked via equipercentile linking analysis (see Kolen & Brennan, 2004) to scores on the state test. This process creates tables of approximately equivalent scores on each assessment, allowing for the lookup of Star scale scores that correspond to the cut scores for different performance levels on the state test. For example, if 20% of students were "Below Basic" on the state test, the lowest Star cut score would be set at a score that partitioned only the lowest 20% of scores.

School-Level Data

While using student-level data is still common, obstacles associated with individual data often lead to a difficult and time-consuming process of obtaining and analyzing data. In light of the time-sensitive needs of schools, obtaining student-level data is not always an option. As an alternative, school-level data may be used in a similar manner. These data are publicly available, thus making the linking process more efficient.

School-level data were analyzed for some of the states included in the student-level linking analysis. In an effort to increase sample size, the school-level data presented here represent "projected" Scaled Scores. Each Star score was projected to the mid-point of the state test administrations window using decile-based growth norms. The growth norms are both grade- and subject-specific and are based on the growth patterns of more than one million students using Star assessments over a three-year period. Again, the linking process used for school-level data is very similar to the previously described process—the distribution of state test scores is compared to projected Star scores and using the observed distribution of state-test scores, equivalent cut scores are created for the Star assessments (the key difference being that these comparisons are made at the group level).

Accuracy Comparisons

Accuracy comparisons between student- and school-level data are particularly important given the marked resource differences between the two methods. These comparisons are presented for three states[1] in Table 58, Table 59, and Table 60. With few exceptions, results of linking using school-level data were nearly

1. Data were available for Arkansas, Florida, Idaho, Kansas, Kentucky, Mississippi, North Carolina, South Dakota, and Wisconsin; however, only North Carolina, Mississippi, and Kentucky are included in the current analysis.

identical to student-level data on measures of specificity, sensitivity, and overall accuracy. McLaughlin and Bandeira de Mello (2002) employed similar methods in their comparison of NAEP scores and state assessment results, and this method has been used several times since then (McLaughlin & Bandeira de Mello, 2003; Bandeira de Mello, Blankenship, & McLaughlin, 2009; Bandeira et al., 2008).

In a similar comparison study using group-level data, Cronin et al. (2007) observed cut score estimates comparable to those requiring student-level data.

Table 58: Number of Students Included in Student-Level and School-Level Linking Analyses by State, Grade, and Subject

State	Grade	Math	
		Student	School
NC	3	1,100	524
	4	751	890
	5	482	551
	6	202	515
	7	216	67
	8	39	372
MS	3	1,786	4,309
	4	1,757	4,584
	5	1,531	5,294
	6	1,180	5,190
	7	721	3,390
	8	549	1,896
KY	3	3,777	935
	4	3,155	1,797
	5	2,228	1,430
	6	1,785	1,497
	7	788	984
	8	362	1,036

Table 59: Comparison of School Level and Student Level Classification Diagnostics for Mathematics

State	Grade	Sensitivity[a]		Specificity[b]		False + Rate[c]		False – Rate[d]		Overall Rate	
		Student	School	Student	School	Student	School	Student	School	Student	School
NC	3	92%	81%	53%	73%	47%	27%	8%	19%	80%	78%
	4	90%	78%	52%	73%	48%	27%	10%	22%	80%	78%
	5	83%	83%	62%	57%	38%	43%	17%	17%	75%	74%
	6	94%	87%	42%	65%	58%	35%	6%	13%	74%	83%
	7	91%	88%	61%	69%	39%	31%	9%	12%	81%	84%
	8	89%	77%	58%	76%	42%	24%	11%	23%	77%	77%
MS	3	78%	70%	77%	83%	23%	17%	22%	30%	77%	76%
	4	73%	73%	81%	81%	19%	19%	27%	27%	77%	77%
	5	71%	68%	83%	84%	17%	16%	29%	32%	77%	76%
	6	71%	66%	81%	85%	19%	15%	29%	34%	76%	76%
	7	83%	84%	82%	81%	18%	19%	17%	16%	83%	83%
	8	56%	66%	89%	83%	11%	17%	44%	34%	76%	76%
KY	3	95%	92%	45%	54%	55%	46%	5%	8%	83%	83%
	4	92%	87%	47%	60%	53%	40%	8%	13%	80%	80%
	5	90%	90%	51%	50%	49%	50%	10%	10%	77%	77%
	6	82%	80%	64%	68%	36%	32%	18%	20%	75%	75%
	7	72%	68%	81%	85%	19%	15%	28%	32%	76%	76%
	8	59%	66%	89%	85%	11%	15%	41%	34%	74%	76%

a. Sensitivity refers to the proportion of correct positive predictions.
b. Specificity refers to the proportion of negatives that are correctly identified (e.g. student will not meet a particular cut score)
c. False + rate refers to the proportion of students incorrectly identified as “at-risk.”
d. False – rate refers to the proportion of students incorrectly identified as not “at-risk.”

Table 60: Comparison of Differences Between Achieved ad Forecasted Performance Levels in Math (Forecast % – Achieved %)

State	Grade	Student	School	Student	School	Student	School	Student	School
NC		Level I		Level II		Level III		Level IV	
	3	–2.6%	–1.6%	–2.8%	0.80%	15.60%	2.10%	–10.2%	–1.3%
	4	–4.0%	–0.4%	–2.5%	1.20%	14.70%	1.50%	–8.2%	–2.3%
	5	–2.7%	–0.9%	1.60%	–3.9%	10.00%	11.60%	–8.9%	–6.7%
	6	–7.3%	–5.3%	–8.2%	–4.5%	18.60%	7.10%	–3.1%	2.70%
	7	–1.3%	–0.6%	–5.0%	–1.1%	15.10%	1.10%	–8.8%	0.60%
	8	–4.2%	–4.4%	–5.6%	–2.9%	2.50%	–1.2%	7.40%	8.60%
MS		Minimal		Basic		Proficient		Advanced	
	3	2.70%	10.10%	0.00%	0.20%	1.10%	–15.0%	–3.9%	4.60%
	4	1.50%	9.90%	4.40%	–3.4%	–3.7%	–10.7%	–2.1%	4.20%
	5	0.80%	9.40%	5.30%	–1.0%	–3.5%	–11.3%	–2.7%	2.80%
	6	4.70%	12.60%	–0.8%	–4.3%	–1.8%	–11.6%	–2.1%	3.30%
	7	0.70%	2.80%	–0.5%	–3.7%	0.00%	–1.8%	–0.2%	2.80%
	8	5.80%	7.00%	4.60%	–4.4%	–9.9%	–4.1%	–0.5%	1.50%
KY		Novice		Apprentice		Proficient		Distinguished	
	3	–3.2%	–2.0%	–4.8%	–2.6%	12.10%	3.30%	–4.0%	1.40%
	4	–4.1%	–2.7%	–3.9%	1.00%	5.60%	1.60%	2.40%	0.10%
	5	–3.7%	–0.2%	–5.4%	–9.7%	11.40%	8.40%	–2.3%	1.60%
	6	–3.9%	–0.4%	0.10%	–0.5%	5.80%	0.50%	–2.1%	0.20%
	7	–1.9%	7.10%	10.50%	3.60%	1.20%	–3.0%	–9.6%	–7.5%
	8	1.50%	4.30%	13.80%	4.90%	–5.0%	–1.9%	–10.2%	–7.3%

References

Allington, R., & McGill-Franzen, A. (2003). Use students' summer-setback months to raise minority achievement. *Education Digest,* 69(3), 19–24.

Alonzo, A. C., & Gearhart, M. (2006). Considering learning progressions from a classroom assessment perspective. *Measurement, 14*(1 & 2), 99–104.

Bandeira de Mello, V., Blankenship, C., & McLaughlin, D. H. (2009). *Mapping state proficiency standards onto NAEP Scales: 2005–2007* (NCES 2010-456). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4), 42–51.

Betebenner, D. W. (2010). New directions for student growth models. Retrieved from the National Network of State Teachers of the Year (NNSTOY) website: https://www.nnstoy.org/download/evalution/LinkClick%20Betebenner.pdf

Betebenner, D. W., & Iwaarden, A. V. (2011a). SGP: An R package for the calculation and visualization of student growth percentiles [Computer Software manual]. (R package version 0.4-0.0 available at http://cran.r-project.org/web/packages/SGP/)

Betebenner, D. W. (2011b). A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth trajectories/projections. The National Center for the Improvement of Educational Assessment. Retrieved from http://www.nj.gov/education /njsmart/performance/SGP_Technical_Overview.pdf

Bock, R. D., Thissen, D., & Zimowski, M. F. (1997). IRT estimation of domain scores. *Journal of Educational Measurement, 34,* 197–211.

Bracey, G. (2002). Summer loss: The phenomenon no one wants to deal with. *Phi Delta Kappan, 84*(1), 12–13.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform of teaching.* (CPRE Research Report # RR-63). New York: Consortium for Policy Research.

Cormier, D. & Bulut, O. (Manuscript under review.) Using computer-adaptive tests to monitor student progress: optimal slope estimates, administration frequency, and data points.

Cronin, J., Kingsbury, G. G., Dahlin, M., & Bowe, B. (2007). *Alternate methodologies for estimating state standards on a widely used computer adaptive test.* Paper presented at the American Educational Research Association, Chicago, IL.

Deno, S. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*(3), 184–192.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93–99.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment.* Washington, DC: Chief Council of State School Officers.

Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling,* 6, 1–55.

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I Error and Power Rates Using an Effect Size Measure with the Logistic Regression Procedure for DIF Detection. *Applied Measurement in Education,* 14, 329–349.

Kolen, M., & Brennan, R. (2004). *Test equating, scaling, and linking* (2nd ed.). New York: Springer.

Leahy, S., & Wiliam, D. (2011). *Devising learning progressions assessment.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

McLaughlin, D. H., & Bandeira de Mello, V. (2002). *Comparison of state elementary school mathematics achievement standards using NAEP 2000.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

McLaughlin, D. H., & Bandeira de Mello, V. (2003). *Comparing state reading and math performance standards using NAEP.* Paper presented at the National Conference on Large-Scale Assessment, San Antonio, TX.

McLaughlin, D. H., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H., et al. (2008). *Comparison between NAEP and state* *reading assessment results: 2003* (NCES 2008-474). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Market Data Retrieval. (2001). A D&B Company: Shelton, CT.

MetaMetrics, Inc. (2004). *The Quantile Framework for Mathematics: Overview*.Durham, NC: Author.

MetaMetrics, Inc. (2019). *Linking the Renaissance Star Assessments™ for Math with the Quantile Framework® for Mathematics.* Durham, NC: Author, February 2019.

MetaMetrics, Inc. (2015). *The Quantile Framework for Mathematics: Quantile Development and Validation.* Durham, NC: Author.

Monaghan, W. (2006). The facts about subscores (ETS R&D Connections No. 4). Princeton, NJ: Educational Testing Services. Available online: http://www.ets.org/Media/Research/pdf/RD_Connections4.pdf.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika,* 78, 691–692.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics.* Reston, VA.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education. Available online: http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). New York: WCB McGraw-Hill.

Owen, R. J. (1969) A Bayesian approach to tailored testing. *Research Bulletin 69-92*. Princeton, N. J.: Educational Testing Service.

Owen, R. J. (1975) A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association,* 70, 351–356.

Pepe M. S., Janes, H., Longton, G. Leisenring, W., & Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology, 159,* 882–890.

Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989). Scaling, Norming, and Equating. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed. pp.

221– 262). New York: American Council on Education and Macmillan Publishing Company.

Popham, W. J. (2007). The lowdown on learning progressions. *Educational Leadership, 64*(7), 83–84.

Rogers, H. J., & Swaminathan, H. (1993). A comparison of the logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement,* 17, 105–116.

Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. *Practical Assessment Research & Evaluation, 7(14)*.

Rudner, L. M. (2005). Expected classification accuracy. *Practical Assessment Research & Evaluation, 10(13)*.

Skoupski, W.P., & Carvajal, J. (2010). A comparison of approaches for improving the reliability of objective scores. *Educational and Psychological Measurement, 70*(3), 357–375.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement, 27, 361–370.

Swaminathan, H. (1994). Differential item functioning: A discussion. In Dany Laveault, Bruno D. Zumbo, Marc E. Gessaroli, and Marvin W. Boss (Eds.), *Modern Theories of Measurement: Problems and Issues.* Ottawa, Canada: University of Ottawa.

Urry, V. W. (1975). *The effects of guessing on parameters of item discriminatory power.* TN 75-2. Washington DC: Personnel Research and Development Center, US Civil Service Commission, May.

VanDerHeyden, A., & Burns, M. (2008). Examination of the Utility of Various Measures of Mathematics Proficiency. *Assessment for Effective Intervention, 33*(4), 215–224.

Wyse, A. E., & Hao, S. (2012). An evaluation of item response theory classification accuracy and consistency indices. *Applied Psychological Measurement, 36(7)*, 602–624.

Zhou, X. H., Obuchowski, N. A., & Obushcowski, D. M. (2002). *Statistical methods in diagnostic medicine.* Wiley & Sons: New York.

Zumbo, B. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4,* 223–233.

Index

Symbols

A

B

C

D

N

O

P

Q

R

S

T

U

V

W

About Renaissance

Renaissance® transforms data about how students learn into instruments of empowerment for classroom teachers, enabling them to guide all students to achieve their full potentials. Through smart, data-driven educational technology solutions that amplify teachers' effectiveness, Renaissance helps teachers teach better, students learn better, and school administrators lead better. By supporting teachers in the classroom but not supplanting them, Renaissance solutions deliver tight learning feedback loops: between teachers and students, between assessment of skills mastery and the resources to spur progress, and between each student's current skills and future academic growth.

Renaissance

R54119.230825

Star Assessments™ for Reading Technical Manual

Renaissance Learning
PO Box 8036
Wisconsin Rapids, WI 54495-8036
Telephone: (800) 338-4204
(715) 424-3636
Outside the US: 1.715.424.3636
Fax: (715) 424-4242
Email (general/technical questions): support@renaissance.com
Email (international support): worldsupport@renaissance.com
Website: www.renaissance.com

Copyright Notice

8/2023 SRRP

Contents

Introduction

Star Reading: Screening and Progress-Monitoring Assessment

Since the 2011–2012 school year, two different versions of Star Reading have been available for use in assessing the reading achievement of students in grades K–12. The comprehensive version is a 34-item standards-based adaptive assessment, aligned to state and national curriculum standards, that takes an average of less than 20 minutes. A shorter, 25-item version assesses reading comprehension only, and takes an average of less than 10 minutes, making it a popular choice for progress monitoring in programs such as Response to Intervention. Both versions provide immediate feedback to teachers and administrators on each student's reading development.

Star Reading Purpose

As a periodic progress-monitoring assessment, Star Reading progress monitoring serves three purposes for students with at least 100-word sight vocabulary. First, it provides educators with quick and accurate estimates of reading comprehension using students' instructional reading levels. Second, it assesses reading achievement relative to national norms. Third, it provides the means for tracking growth in a consistent manner longitudinally for all students. This is especially helpful to school- and district-level administrators.

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

The Star Reading test's repeatability and flexible administration provide specific advantages for everyone responsible for the education process:

- For students, Star Reading software provides a challenging, interactive, and brief test that builds confidence in their reading ability.
- For teachers, the Star Reading test facilitates individualized instruction by identifying children who need remediation or enrichment most.
- For principals, the Star Reading software provides regular, accurate reports on performance at the class, grade, building, and district level.
- For district administrators and assessment specialists, it provides a wealth of reliable and timely data on reading growth at each school and districtwide. It also provides a valid basis for comparing data across schools, grades, and special student populations.

This manual documents the suitability of Star Reading computer-adaptive testing for these purposes and demonstrates quantitatively how well this innovative instrument in reading assessment performs.

Star Reading is similar in many ways to the Star Reading Progress Monitoring version, but with some enhanced features, including additional reports and expanded benchmark management.

Design of Star Reading

Three Generations of Star Reading Assessments

The introduction of the current version of Star Reading in 2011 marked the third generation of Star Reading assessments. The first generation consisted of Star Reading version 1, which was a variable-length adaptive assessment of reading comprehension that employed a single item type: vocabulary-in-context items. Star Reading's original item bank contained 800+ such items. Although it was a breakthrough computer adaptive test, Star Reading 1 was based on classical test theory.

The second generation consisted of Star Reading versions 2 through 4.4, including the current Star Reading Progress Monitoring version. This second generation differed from the first in three major respects: It replaced classical test theory with Item Response Theory (IRT) as the psychometric foundation for adaptive item selection and scoring; its test length was fixed at twenty-five items (rather than the variable length of version 1); and its content included a second item type: the original vocabulary in context items were augmented in grades 3–12 by the use of longer, authentic text passages for the last 5 items of each test. The second generation versions differed from one another primarily in terms of the size of their item banks, which grew to over 2000 items in version 4.4. Like the first generation of Star Reading tests, the second generation continued to measure a single construct: reading comprehension.

The third generation is represented by the current version of Star Reading. This is the first version of Star Reading to be designed as a standards-based test; its items are organized into 5 blueprint domains, 10 skill sets, 36 general skills, and over 470 discrete skills—all designed to align to national and state curriculum standards in reading and language arts, including the Common Core State Standards. Like the second generation of Star Reading tests, the third generation Star uses fixed-length adaptive tests. Its tests are longer than the second generation test—34 items in length—both to facilitate broader standards coverage and to improve measurement precision and reliability.

Overarching Design Considerations

One of the fundamental Star Reading design decisions involved the choice of how to administer the test. The primary advantage of using computer software to administer Star Reading tests is the ability to tailor each student's test based on his or her responses to previous items. Conventional assessments, including paper-and-pencil tests, typically entail fixed test forms: every student must respond to the same items in the same sequence. Using computer-adaptive procedures, it is possible for students to test on items that appropriately match their current level of proficiency. The item selection procedures, termed Adaptive Branching, effectively customize the test for each student's achievement level.

Adaptive Branching offers significant advantages in terms of test reliability, testing time, and student motivation. Reliability improves over fixed-form tests because the test difficulty is adjusted to each individual's performance level; students do not have to fit a "one test fits all" model. Most of the test items that students respond to are at levels of difficulty that closely match their achievement level. Testing time decreases because, unlike in paper-and-pencil tests, there is no need to expose every student to a broad range of material, portions of which are inappropriate because they are either too easy for high achievers or too difficult for those with low current levels of performance. Finally, student motivation improves simply because of these issues—test time is minimized and test content is neither too difficult nor too easy.

Another fundamental Star Reading design decision involved the choice of the content and format of items for the test. Many types of stimulus and response procedures were explored, researched, discussed, and prototyped. These procedures included the traditional reading passage followed by sets of literal or inferential questions, previously published extended selections of text followed by open-ended questions requiring student-constructed answers, and several cloze-type procedures for passage presentation. While all of these procedures can be used to measure reading comprehension and overall reading achievement, the vocabulary-in-context format was selected as the primary item format for the first generation Star Reading assessments. This decision was made for interrelated reasons of efficiency, breadth of construct coverage, and objectivity and simplicity of scoring.

Four fundamental arguments support the use of the original Star Reading design for obtaining quick and reliable estimates of reading comprehension and reading achievement:

1. The vocabulary-in-context test items, while using a common format for assessing reading, require reading comprehension. Each test item is a complete, contextual sentence with a tightly controlled vocabulary level. The semantics and syntax of each context sentence are arranged to provide clues as to the correct cloze word. The student must actually interpret the meaning of (in other words, comprehend) the sentence in order to choose the correct answer because all of the answer choices “fit” the context sentence either semantically or syntactically. In effect, each sentence provides a mini-selection on which the student demonstrates the ability to interpret the correct meaning. This is, after all, what most reading theorists believe reading comprehension to be—the ability to draw meaning from text.
2. In the course of taking the vocabulary-in-context section of Star Reading tests, students read and respond to a significant amount of text. The Star Reading test typically asks the student to demonstrate comprehension of material that ranges over several grade levels. Students will read, use context clues from, interpret the meaning of, and attempt to answer 20 to 25 cloze sentences across these levels, generally totaling more than 300 words. The student must select the correct word from sets of words that are all at the same reading level, and that at least partially fit the sentence context. Students clearly must demonstrate reading comprehension to correctly respond to these 20 to 25 questions.
3. A child’s level of vocabulary development is a major factor—perhaps *the* major factor—in determining his or her ability to comprehend written material. Decades of reading research have consistently demonstrated that a student’s level of vocabulary knowledge is the most important single element in determining the child’s ability to read with comprehension. Tests of vocabulary knowledge typically correlate better than do any other components of reading with valid assessments of reading comprehension. In fact, vocabulary tests often relate more closely to sound measures of reading comprehension than various measures of comprehension do to each other. Knowledge of word meaning is simply a fundamental component of reading comprehension.

4. The student's performance on the vocabulary-in-context section is used to determine the initial difficulty level of the subsequent authentic text passage items. Although this section consists of just five items, the accurate entry level and the continuing adaptive selection process mean that all of the authentic text passage items are closely matched to the student's reading ability level. This results in unusually high measurement efficiency.

The current third-generation tests expand the breadth of item formats and content beyond that of the previous versions. Each test consists of 34 items; of these, the first 10 are vocabulary-in-context items, while the last 24 items spiral their content to include standards-based material from all five blueprint domains.

The introduction of the 34-Item Star Reading version does not replace the previous version or make it obsolete. The previous version continues to be available as "Star Reading Progress Monitoring," the familiar 25-item measure of reading comprehension. Star Reading thus gives users a choice between a brief assessment focusing on reading comprehension alone, or a longer, standards-based assessment which assures that a broad range of different reading skills, appropriate to student grade level and performance, are included in each assessment.

For these reasons, the Star Reading test design and item format provide a valid procedure for assessing a student's reading comprehension. Data and information presented in this manual reinforce this.

Improvements Specific to Star Reading Versions 3 and Higher

Versions 3 and 4 are adaptations of version 2 designed specifically for use on a computer with web access. In versions 3 and higher, all management and test administration functions are controlled using a management system which is accessed by means of a computer with web access.

This makes a number of new features possible:

- It makes it possible for multiple schools to share a central database, such as a district-level database. Records of students transferring between schools within the district will be maintained in the database; the only information that needs revision following a transfer is the student's updated school and class assignments.
- The same database that contains Star Reading data can contain data on other Star tests, including Star Early Literacy and Star Math. The Renaissance program is a powerful information management program that allows you to manage all your district, school, personnel, parent, and student data in one place. Changes made to district, school, teacher, parent, and student data for

any of these products, as well as other Renaissance software, are reflected in every other Renaissance program sharing the central database.

- Multiple levels of access are available, from the test administrator within a school or classroom to teachers, principals, district administrators, and even parents.
- Renaissance takes reporting to a new level. Not only can you generate reports from the student level all the way up to the school level, but you can also limit reports to specific groups, subgroups, and combinations of subgroups. This supports "disaggregated" reporting; for example, a report might be specific to students eligible for free or reduced lunch, to English language learners, or to students who fit both categories. It also supports compiling reports by teacher, class, school, grade within a school, and many other criteria such as a specific date range. In addition, the Renaissance consolidated reports allow you to gather data from more than one program (such as Star Reading and Accelerated Reader) at the teacher, class, school, and district level and display the information in one report.
- Since the Renaissance software is accessed through a web browser, teachers (and administrators) will be able to access the program from home—provided the district or school gives them that access.

Test Interface

The Star Reading test interface was designed to be both simple and effective. Students can use either the mouse or the keyboard to answer questions.

- If using the keyboard, students press one of the four number keys (**1**, **2**, **3**, and **4**) and then press the **Enter** key (or the **return** key on Macintosh computers).
- If using the mouse, students click the answer of choice and then click **Next** to enter the answer.
- On a tablet, students tap their answer choice; then, they tap **Next**.

Practice Session

Star Reading software includes a provision for a brief practice test preceding the test itself. The practice session allows students to get comfortable with the test interface and to make sure that they know how to operate it properly. As soon as a student has answered three practice questions correctly, the program takes the student into the actual test. As long as they possess the requisite 100-word vocabulary, even the lowest-level readers should be able to answer the sample questions correctly. If the student has not successfully answered three items by the end of the practice session, Star Reading will halt the testing session and tell the student to ask the teacher for help. It may be that the student cannot read at even the most basic level, or it may be that the student needs help operating the interface, in which case the teacher should help the student through the practice session the next time. Before beginning the next test with the student, the program will recommend that the teacher assist the student during the practice.

Once a student has successfully passed a practice session, the student will not be presented with practice items again on a test of the same type taken within the next 180 days.

Adaptive Branching/Test Length

Star Reading's branching control uses a proprietary approach somewhat more complex than the simple Rasch maximum information IRT model. The Star Reading approach was designed to yield reliable test results for both the criterion-referenced and norm-referenced scores by adjusting item difficulty to the responses of the individual being tested while striving to minimize test length and student frustration.

In order to minimize student frustration, the first administration of the Star Reading test begins with items that have a difficulty level that is below what a typical student at a given grade can handle—usually one or two grades below grade placement. On the average, about 85 percent of students will be able to answer the first item correctly. Teachers can override this typical value by entering an even lower Estimated Instructional Reading Level for the student. On the second and subsequent administrations, the Star Reading test again begins with items that have a difficulty level lower than the previously demonstrated reading ability. Students generally have an 85 percent chance of answering the first item correctly on second and subsequent tests.

Test Length

Once the testing session is underway, the Star Reading test administers 34 items (the Star Reading Progress Monitoring test administers 25 items) of varying difficulty based on the student's responses; this is sufficient information to obtain a reliable Scaled Score and to determine the student's Instructional Reading Level.

The length of time needed to complete a Star Reading test varies across students.

Table 1 provides an overview of the testing time by grade for the students who took the full-length 34-item version of Star Reading during the 2018–2019 school year. The results of the analysis of test completion time indicate that half or more of students completed the test in less than 20 minutes, depending on grade, and even in the slowest grade (grade K) 95% of students finished their Star Reading test in less than 34 minutes.

Table 2 provides an overview of the Star Reading Progress Monitoring testing time by grade for the students using data from the 2017–2018 and 2018–2019 school years. For that version of the test, about half of the students at every grade completed the Star Reading Progress Monitoring test in less than 10 minutes, and even in the slowest grade (grade 1) 95 percent of students finished in less than 18 minutes.

Table 1: Average and Percentiles of Total Time to Complete the 34-item Star Reading Assessment During the 2018–2019 School Year

Grade	Sample Size	Time to Complete Test (in Minutes)					
		Mean	Standard Deviation	5th Percentile	50th Percentile	95th Percentile	99th Percentile
K	77,319	18.39	8.16	8.73	16.68	34.00	43.83
1	1,734,368	18.84	7.38	9.08	17.75	32.50	40.52
2	3,574,122	19.17	6.57	9.80	18.42	31.08	37.63
3	4,047,336	18.78	5.22	10.47	18.55	27.75	31.60
4	3,872,024	19.75	5.21	11.18	19.65	28.55	32.12
5	3,758,949	19.63	5.01	11.42	19.53	28.07	31.60
6	2,827,076	19.59	4.89	11.48	19.53	27.75	31.13
7	2,190,539	19.33	4.83	11.25	19.30	27.35	30.67
8	2,063,913	19.13	4.80	11.12	19.12	27.10	30.45
9	914,315	18.92	4.87	10.68	18.93	26.95	30.27
10	724,030	18.51	4.9o	10.35	18.48	26.67	30.12
11	448,315	18.25	4.98	10.02	18.22	26.55	30.03
12	275,495	17.95	5.12	9.70	17.85	26.58	30.18

Table 2: Average and Percentiles of Total Time to Complete the 25-item Star Reading Progress Monitoring Assessment During the 2017–2018 and 2018–2019 School Years

Grade	Sample Size	Time to Complete Test (in Minutes)					
		Mean	Standard Deviation	5th Percentile	50th Percentile	95th Percentile	99th Percentile
1	10,260	10.27	3.55	5.62	9.67	17.15	20.37
2	31,898	9.35	2.87	5.60	8.83	14.85	17.75
3	33,128	9.67	2.57	5.95	9.38	14.43	16.67
4	31,340	9.48	2.46	5.93	9.20	13.98	16.33
5	28,656	9.35	2.47	5.82	9.03	13.93	16.00
6	14,980	9.02	2.42	5.65	8.68	13.52	16.07
7	10,196	8.71	2.29	5.57	8.40	12.95	15.10
8	10,232	8.59	2.33	5.45	8.25	12.93	15.67
9	1,800	8.55	2.34	5.45	8.13	13.14	15.31
10	1,451	8.11	2.07	5.32	7.78	11.85	14.12
11	738	8.00	2.10	5.32	7.62	12.18	14.12
12	483	7.92	2.09	5.30	7.67	11.97	14.93

Test Repetition

Star Reading score data can be used for multiple purposes such as screening, placement, planning instruction, benchmarking, and outcomes measurement. The frequency with which the assessment is administered depends on the purpose for assessment and how the data will be used. Renaissance Learning recommends assessing students only as frequently as necessary to get the data needed. Schools that use Star for screening purposes typically administer it two to five times per year. Teachers who want to monitor student progress more closely or use the data for instructional planning may use it more frequently. Star Reading may be administered monthly for progress monitoring purposes, and as often as weekly when needed.

Star Reading keeps track of the questions presented to each student from test session to test session and will not ask the same question more than once in any 120-day period.

Item Time Limits

Star Reading tests place no limits on total testing time. However, there are time limits for each test item. The per-item time limits are generous, and ensure that more than 90 percent of students can complete each item within the normal time limits.

Star Reading provides the option of extended time limits for selected students who, in the judgment of the test administrator, require more than the standard amount of time to read and answer the test questions.

Extended time may be a valuable accommodation for English language learners as well as for some students with disabilities. Test users who elect the extended time limit for their students should be aware that Star Reading norms, as well as other technical data such as reliability and validity, are based on test administration using the standard time limits. When the extended time limit accommodation is elected, students have three times longer than the standard time limits to answer each question.

Table 3 shows the Star Reading Progress Monitoring version's test time-out limits for individual items. These time limits are based on a student's grade level.

Table 3: Star Reading Progress Monitoring Time-Out Limits

Grade	Question Type	Standard Time Limit (seconds/ item)	Extended Time Limit (seconds/item)
K–2	Practice	60	180
	Test, questions 1–25[a]	60	180
	Skill Test—Practice (Calibration)	60	180
	Skill Test—Test (Calibration)	60	180
3–12	Practice	60	180
	Test, questions 1–20[a]	45	135
	Test, questions 21–25[b]	90	270
	Skill Test—Practice (Calibration)	60	180
	Skill Test—Test (Calibration)	90	270

a. Vocabulary-in-context items.
b. Authentic text/passage comprehension items.

These time-out values are based on latency data obtained during item validation. Very few vocabulary-in-context items at any grade had latencies longer than 30 seconds, and almost none (fewer than 0.3 percent) had latencies of more than 45 seconds. Thus, the time-out limit was set to 45 seconds for most students and increased to 60 seconds for the very young students. Longer time limits were allowed for the lengthier authentic text passages items.

Table 4 shows time limits for the 34-item Star Reading version's test questions:

Table 4: Star Reading Time-Out Limits

Grade	Question Type	Standard Time Limit (seconds/ item)	Extended Time Limit (seconds/ item)
K–2	Practice	60	180
	Test Section A, questions 1–10[a]	120	360
	Test Section B, questions 11–34[b]	180	405
3–12	Practice	60	180
	Test Section A, questions 1–10[a]	105	315
	Test Section B, questions 11–34[b]	150	450

a. Vocabulary-in-context items.
b. Items from 5 domains in 5 blocks, including some vocabulary-in-context.

At all grades, regardless of the extended time limit setting, when a student has only 15 seconds remaining for a given item, a time-out warning appears, indicating that he or she should make a final selection and move on. Items that time out are counted as incorrect responses *unless* the student has the correct answer selected when the item times out. If the correct answer is selected at that time, the item will be counted as a correct response.

If a student doesn't respond to an item, the item times out and briefly gives the student a message describing what has happened. Then the next item is presented. The student does not have an opportunity to take the item again. If a student doesn't respond to any item, all items are scored as incorrect.

Accessibility and Test Accommodations

The Star Reading test can be accessed in an accessible format that is in compliance with WCAG 2.1 AA. This format allows for users with different ability levels to access the test utilizing different modalities, including assistive technology such as the JAWS screen reader. The content of the item bank is the same as the traditional item delivery format, although the user interface is modified slightly. A student will be presented with the WCAG 2.0 AA version of the test after educators select one of the relevant test accommodations available in that student's Personal Needs Profile. Some of the available accommodations are the ability to change the size of the text or the color contrast, a highlighter, a line reader, an answer choice eliminator or unlimited time to answer questions. In order to provide the best experience for students and teachers, the available accommodations could be modified during the school year.

Unlimited Time

Beginning with the 2022–23 school year, a new preference has been added: the Accommodations Preference. Among other things, this preference allows teachers to give students virtually unlimited time to answer questions: 15 minutes for both practice questions and test questions. When this preference is set, the student will not see a time-out warning when there are 15 seconds left; however, if there is no activity at all from the student within 15 minutes of a question first being presented, the student will be shown a dialog box. The student will have 60 seconds to close the dialog box and return to the test. If the student does not close the dialog box within 60 seconds, the student’s current progress on the test will be saved and the test will be ended (and can be resumed the same way as a paused test).

Test Security

Star Reading software includes a number of security features to protect the content of the test and to maintain the confidentiality of the test results.

Split-Application Model

When students log into Star Reading, they do not have access to the same functions that teachers, administrators, and other personnel can access. Students are allowed to take the test, but no other features available in Star Reading are available to them; therefore, they have no access to confidential information. When teachers and administrators log in, they can manage student and class information, set preferences, and create informative reports about student test performance.

Individualized Tests

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

Data Encryption

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

Access Levels and Capabilities

Each user's level of access to a Renaissance program depends on the primary position assigned to that user. Each primary position is part of a user permission group. There are six of these groups: district level administrator, district dashboard owner, district staff, school level administrator, school staff, and teacher. By default, each user permission group is granted a specific set of user permissions; each user permission corresponds to one or more tasks that can be performed in the program. The user permissions for these groups can be changed, and user permissions can be granted or removed on an individual level.

Renaissance also allows you to restrict students' access to certain computers. This prevents students from taking Star Reading tests from unauthorized computers (such as home computers). For more information, see https://help.renaissance.com/setup/22509.

The security of the Star Reading data is also protected by each person's user name (which must be unique) and password. User names and passwords identify users, and the program only allows them access to the data and features that they are allowed based on their primary position and the user permissions that they have been granted. Personnel who log in to Renaissance (teachers, administrators, or staff) must enter a user name and password before they can access the data and create reports. Parents who are granted access to Renaissance must also log in with a user name and password before they can access information about their children. Without an appropriate user name and password, personnel and parents cannot use the Star Reading software.

Test Monitoring/Password Entry

Test monitoring is another useful Star Reading security feature. Test monitoring is implemented using the Password Requirement preference, which specifies whether monitors must enter their passwords at the start of a test. Students are required to enter a user name and password to log in before taking a test. This ensures that students cannot take tests using other students' names.

Final Caveat

While Star Reading software can do much to provide specific measures of test security, the most important line of defense against unauthorized access or misuse of the program is the user's responsibility. Teachers and test monitors need to be careful not to leave the program running unattended and to monitor all testing to prevent students from cheating, copying down questions and answers, or performing "print screens" during a test session. Taking these simple precautionary steps will help maintain Star Reading's security and the quality and validity of its scores.

Test Administration Procedures

In order to ensure consistency and comparability of results to the Star Reading norms, students taking Star Reading tests should follow standard administration procedures. The testing environment should be as free from distractions for the student as possible.

The Test Administration Manual included with the Star Reading product describes the standard test orientation procedures that teachers should follow to prepare their students for the Star Reading test. These instructions are intended for use with students of all ages; however, the Star Reading test should only be administered to students who have a reading vocabulary of at least 100 words. The instructions were successfully field-tested with students ranging from grades 1–8. It is important to use these same instructions with all students before they take the Star Reading test.

Content and Item Development

Content Specification: Star Reading

The scale and scope of Star Reading content has steadily grown since content development first started many years ago. Since the original test in 1995, which was exclusively Vocabulary-in-Context, other item types have been added to test additional skills, and new item designs continue to be added as state standards evolve.

Star Reading is based upon the assessment of 36 Blueprint Skills organized within 5 Blueprint Domains of reading (see Table 5), and maps the progressions of reading skills and understandings as they develop in sophistication from kindergarten through grade 12. Each Star item is designed to assess a specific skill within the test blueprint. The test blueprint is structured to provide a consistent assessment experience even as state-specific Renaissance Reading Learning Progressions may change, as well as the set of items associated with the blueprint. The Star Reading test blueprint is largely fixed. Renaissance may alter the blueprint if there are data-driven reasons to make a major change to the content.

For information regarding the development of Star Reading items, see "Item Development Specifications: Star Reading" on page 19. Before inclusion in the Star Reading item bank, all items are reviewed to ensure they meet the content specifications for Star Reading item development. Items that do not meet the specifications are either discarded or revised for recalibration. All new item development adheres to the content specifications and all items have been calibrated using the dynamic calibration method.

The first stage of expanded Star Reading development is to identify the set of skills to be assessed. Multiple resources were consulted to determine the set of skills most appropriate for assessing the reading development of K–12 US students. The resources include but are not limited to:

- Reading Next—A Vision for Action and Research in Middle and High School Literacy: A Report to Carnegie Corporation of New York © 2004 by Carnegie Corporation of New York. https://www.all4ed.org/wp-content/uploads/2006/07/ReadingNext.pdf.
- NCTE Principles of Adolescent Literacy Reform, A Policy Research Brief, Produced by The National Council of Teachers of English, April 2006. http://www.ncte.org/library/NCTEFiles/Resources/Positions/Adol-Lit-Brief.pdf.

- Improving Adolescent Literacy: Effective Classroom and Intervention Practices, August 2008. http://eric.ed.gov/PDFS/ED502398.pdf.
- Reading Framework for the 2009 National Assessment of Education Progress. https://www.nagb.gov/assets/documents/publications/frameworks/reading/2009-reading-framework.pdf.
- Common Core State Standards Initiative (2010). Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects.
- Individual state standards from all 50 states.

The development of the skills list included iterative reviews by reading and assessment experts and psychometricians specializing in educational assessment. See Table 5 for the Star Reading Blueprint Skills List. The skills list is organized into five blueprint domains:

- Word Knowledge and Skills
- Comprehension Strategies and Constructing Meaning
- Analyzing Literary Text
- Understanding Author's Craft
- Analyzing Argument and Evaluating Text

The second stage of development includes item development and calibration. Assessment items are developed according to established specifications for grade-level appropriateness and then reviewed to ensure the items meet the specifications. Grade-level appropriateness is determined by multiple factors including reading skill, reading level, cognitive load, vocabulary grade level, sentence structure, sentence length, subject matter, and interest level. All writers and editors have content-area expertise and relevant classroom experience and use those qualifications in determining grade-level appropriateness for subject matter and interest level. A strict development process is maintained to ensure quality item development.

Assessment items, once written, edited, and reviewed, are field tested and calibrated to estimate their Rasch difficulty parameters and goodness of fit to the model. Field testing and calibration are conducted in a single step. This dynamic calibration method is done by embedding new items in appropriate, random positions within the Star assessments to collect the item response data needed for psychometric evaluation and calibration analysis. Following these analyses, each assessment item—along with both traditional and Item Response Theory (IRT) analysis information (including fit plots) and information about the test level, form, and item identifier—is stored in an item statistics database. A panel of content reviewers then examines each item within the proper context, to determine whether the item meets all criteria for use in an operational assessment.

Table 5: Star Reading Assessment Organization: Star Reading Blueprint Domains, Skill Sets, and Skills

Star Reading Blueprint Domain	Star Reading Blueprint Skill Set	Star Reading Blueprint Skill
Word Knowledge and Skills	Vocabulary Strategies	• Use context clues • Use structural analysis
	Vocabulary Knowledge	• Recognize and understand synonyms • Recognize and understand homonyms and multi-meaning words • Recognize connotation and denotation • Understand idioms • Understand analogies
Comprehension Strategies and Constructing Meaning	Reading Process Skills	• Make predictions • Identify author's purpose • Identify and understand text features • Recognize an accurate summary of text
	Constructing Meaning	• Understand vocabulary in context • Draw conclusions • Identify and understand main ideas • Identify details • Extend meaning and form generalizations • Identify and differentiate fact and opinion
	Organizational Structure	• Identify organizational structure • Understand cause and effect • Understand comparison and contrast • Identify and understand sequence
Analyzing Literary Text	Literary Elements	• Identify and understand elements of plot • Identify and understand setting • Identify characters and understand characterization • Identify and understand theme • Identify the narrator and point of view
	Genre Characteristics	• Identify fiction and nonfiction, reality and fantasy • Identify and understand characteristics of genres
Understanding Author's Craft	Author's Choices	• Understand figurative language • Understand literary devices • Identify sensory detail
Analyzing Argument and Evaluating Text	Analysis	• Identify bias and analyze text for logical fallacies • Identify and understand persuasion
	Evaluation	• Evaluate reasoning and support • Evaluate credibility

An Example of Star Reading Item Adherence to a Specific Skill within Star Reading Blueprint Structure

<table>
<tr><td colspan="3">Blueprint Domain: Analyzing literary text</td></tr>
<tr><td colspan="3">Blueprint Skill Set: Literary Elements</td></tr>
<tr><td colspan="3">Blueprint Skill: Identify characters and understand characterization</td></tr>
<tr><td rowspan="3">Grade-level subskill statements:</td><td>2nd grade</td><td>Describe major and minor characters and their traits using key details.</td></tr>
<tr><td rowspan="2">3rd grade</td><td>Identify and describe main characters’ traits, motives, and feelings.</td></tr>
<tr><td>3rd Grade Star Reading Item

Ajay likes being the youngest child in his family. His two older brothers look after him. Before he goes to sleep, they tell him adventure stories. Ajay always falls asleep before the stories are over. The stories will be continued the next night.

How does Ajay feel about his brothers?

1. He wants to get bigger so he can play with them.
2. He likes that they look after him and tell him stories.
3. He wishes their stories didn’t keep him awake.</td></tr>
<tr><td></td><td>4th grade</td><td>Describe characters, interactions with other characters, and relationship between actions, traits, and motives.</td></tr>
</table>

The Educational Development Laboratory’s Core Vocabulary List: ATOS Graded Vocabulary List

The original point of reference for the development of Star Reading items was the 1995 updated vocabulary lists that are based on the Educational Development Laboratory’s (EDL) *A Revised Core Vocabulary* (1969) of 7,200 words. The EDL vocabulary list is a soundly developed, validated list that is often used by developers of educational instruments to create all types of educational materials and assessments. It categorizes hundreds of vocabulary words according to grade placement, from primer (pre-grade 1) through grade 13 (post-high school). This was exactly the span desired for the Star Reading test.

Beginning with new test items introduced in version 4.3, Star Reading item developers have used ATOS instead of the EDL word list. ATOS is a system for evaluating the reading level of continuous text; it contains over 125,000 words in its graded vocabulary list. This readability formula was developed by Renaissance Learning, Inc., and designed by leading readability experts. ATOS is the first formula to include statistics from actual student book reading.

Content Specification: Star Reading

The Content item bank for Star Reading has been expanding steadily since the original product launch and continues to this day. Content development is driven by the test design and test purposes, which are to measure comprehension and general reading achievement. Based on test purpose, the desired content had to meet certain criteria. First, it had to cover a range broad enough to test students from grades K–12. Thus, items had to represent reading levels ranging all the way from kindergarten through post-high school. Second, the current collection of test items must be large enough so that students could test often without being given the same items twice.

The current item bank for Star Reading contains over 6,000 items.

Item Development Specifications: Star Reading

During item development, every effort is made to avoid the use of stereotypes, potentially offensive language or characterizations, and descriptions of people or events that could be construed as being offensive, demeaning, patronizing, or otherwise insensitive. The editing process also includes a strict sensitivity review of all items to attend to issues of gender and ethnic-group balance and fairness.

Vocabulary-in-Context Item Specifications

Each of the vocabulary items is written to the following specifications:

1. Each vocabulary-in-context test item consists of a single-context sentence. This sentence contains a blank indicating a missing word. Three or four possible answers are shown beneath the sentence. For questions developed at a kindergarten or first-grade reading level, three possible answers are given. Questions at a second-grade reading level and higher offer four possible answers.
2. To answer the question, the student selects the word from the answer choices that best completes the sentence. The correct answer option is the word that

appropriately fits both the semantics and the syntax of the sentence. All of the incorrect answer options either fit the syntax of the sentence or relate to the meaning of something in the sentence. They do not, however, meet both conditions.

3. The answer blanks are generally located near the end of the context sentence to minimize the amount of rereading required.

4. The sentence provides sufficient context clues for students to determine the appropriate answer choice. However, the length of each sentence varies according to the guidelines shown in Table 6.

5. Typically, the words that provide the context clues in the sentence are below the level of the actual test word. However, due to a limited number of available words, not all of the questions at or below grade 2 meet this criterion—but even at these levels, no context words are above the grade level of the item.

6. The correct answer option is a word selected from the appropriate grade level of the item set. Incorrect answer choices are words at the same grade level or one grade below. Through vocabulary-in-context test items, Star Reading requires students to rely on background information, apply vocabulary knowledge, and use active strategies to construct meaning from the assessment text. These cognitive tasks are consistent with what researchers and practitioners describe as reading comprehension.

Table 6: Maximum Sentence Length per Item Grade Level

Item Grade Level	Maximum Sentence Length (Including Sentence Blank)
Kindergarten and Grade 1	10 words
Grades 2 and 3	12 words
Grades 4–6	14 words
Grades 7–13	16 words

Authentic Text Passage Item Specifications

Authentic text items are used exclusively as an element of the Star Reading Progress Monitoring test. Authentic text passage items are passages of extended text administered to students at grade levels 3–13. To support students receiving items at grade levels K–3, some original passages were written. Authentic text items were developed by identifying authentic texts, extracting appropriate passages, and creating cloze-type questions and answers. Each passage is comprised of content that can stand alone as a unified, coherent text. Items were selected which assess passage-level, not merely sentence-level, understanding.

To answer the item correctly, the student needs to have a general understanding of the context and content of the passage, not merely an understanding of the specific content of the sentence.

The first authentic passages in Star Reading were extracted from children's and young adult literature, from nonfiction books, and from newspapers, magazines, and encyclopedias. Passages were selected from combinations of three primary categories for school-age children: popular fiction, classic fiction, and nonfiction. Overall Flesch-Kincaid readability estimates of the source materials were used as initial estimates of grade-level difficulty.

After the grade-level difficulty of a passage was estimated, the passage was searched for occurrences of Educational Development Laboratory (EDL) words at the same grade level difficulty. When an EDL word was found that, if replaced with a blank space, would make the passage a good cloze passage, the passage was extracted for use as an authentic text passage test item. Approximately 600 authentic text passage items were initially developed.

Each of the items in the resulting pool was then rated according to several criteria in order to determine which items were best suited for inclusion in the tryout and calibration. Three educators rated each item on the following criteria:

- Grade-level appropriateness of the text
- Cohesiveness of the passage
- Suitability of the passage for its grade level in terms of vocabulary
- Suitability of the passage for its grade level in terms of content density

To ensure a variety of authentic text passage items on the test, each passage was also placed in one of the following categories, according to Meyer and Rice:

1. Antecedent-consequence—causal relationships are found between sentences.
2. Response—a question-answer or a problem-solving format.
3. Comparison—similarities and differences between sentences are found.
4. Collection—sentences are grouped together based on some common idea or event. This would include a sequence of events.
5. Description—sentences provide information by explanation, in specific attributes of the topic, or elaborating on setting.

Replacement passages and newly created items intended for use in versions 4.3 and later were extracted primarily from Accelerated Reader (AR) books. (Updated content specifications were used for writing the new and replacement Star Reading items in version 4.3.) Target words were selected in advance (based on the average ATOS level of target words within a range of difficulty levels). Texts of

AR books, based on those with the fewest quiz requests, were run through a text-analysis tool to find instances of use. This was done to decrease the possibility that students may have already encountered an excerpt.

Consideration was given to include some passages from the public domain. When necessary, original long items were written. In any case, passages excerpted or adapted are attributed in "Item and Scale Calibration" on page 31.

Each of the authentic text passage items is written to the following specifications:

1. Each authentic text passage test item consists of a paragraph. The second half of the paragraph contains a sentence with a blank indicating a missing word. Four possible answers are shown beneath the sentence.
2. To answer the question, the student selects the word from the list of answer choices that best completes the sentence based on the context of the paragraph. The correct answer choice is the word that appropriately fits both the semantics and the syntax of the sentence, and the meaning of the paragraph. All of the incorrect answer choices either fit the syntax of the sentence or relate to the meaning of the paragraph.
3. The paragraph provides sufficient context clues for students to determine the appropriate answer choice. Average sentence length within the paragraphs is 8–16 words depending on the item's grade level. Total passage length ranges from 27–107 words, based on the average reading speed of each grade level, as shown in Table 7.

Table 7: Authentic Text Passage Length

Grade	Average Reading Speed (Words/Minute)	Passage Length (Approximate Number of Words)
1	80	30
2	115	40
3	138	55
4	158	70
5–6	173, 185	80
7–9	195, 204, 214	90
10–12	224, 237, 250	100

4. Answer choices for authentic text passage items are EDL Core Vocabulary or ATOS words selected from vocabulary levels at or below that of the correct response. The correct answer for a passage is a word at the targeted level of the item. Incorrect answers are words or appropriate synonyms at the same EDL or ATOS vocabulary level or one grade below.

Reading Skills Item Specifications

Valid item development is contingent upon several interdependent factors. The following section outlines the factors which guide Star Reading item content development. Item content is comprised of stems, answer choices, and short passages. Additional, detailed information may be found in the English Language Arts Content Appropriateness Guidelines and Item Development Guidelines outlined in the content specification.

Adherence to Skills

Star Reading assesses more than 600 grade-specific skills within the Renaissance Core Progress for Reading Learning Progression. Item development is skill-specific. Each item in the item bank is developed for and clearly aligned to one skill. An item meets the alignment criteria if the knowledge and skill required to correctly answer the item match the intended knowledge and skill being assessed. Answering an item correctly does not require reading skill knowledge beyond the expected knowledge for the skill being assessed. Star Reading items include only the information and text needed to assess the skill.

Level of Difficulty: Readability

Readability is a primary consideration for level of item difficulty. Readability relates to the overall ease of reading a passage and items. Readability involves the reading level, as well as the layout and visual impact of the stem, passage/support information/graphics, and the answer choices. Readability in Star item development accounts for the combined impact, including intensity and density, of each part of the item, even though the individual components of the item may have different readability guidelines.

The reading level and grade level for individual words are determined by ATOS. Item stems and answer choices present several challenges to accurately determining reading level. Items may contain discipline-specific vocabulary that is typically above grade level but may still be appropriate for the item. Examples of this could include *summary, paragraph,* or *organized* and the like. Answer choices may be incomplete sentences for which it is difficult to get an accurate reading grade level. These factors are taken into account when determining reading level.

Item stems and answer choices that are complete sentences are written for the intended grade level of the item. The words in answer choices and stems that are not complete sentences are within the designated grade-level range. Reading comprehension is not complicated by unnecessarily difficult sentence structure and/or vocabulary.

Items and passages are written at grade level. Table 8 indicates the GLE range, item word count range, maximum passage word count range, and sentence length range.

One exception exists for the reading skill *use context clues*. For those items, the target word will be one grade level above the designated grade of the item.

Table 8: Readability Guidelines Table

Grade	GLE Range	Maximum Item Word Count	Sentence Length Range	Number of Words 1 Grade Above (per 100)	Number of Unrecognized Words
K		Less than 30	< 10	0	As a rule, the only unrecognized words will be: names, common derivatives, etc.
1		30	10	0	
2	1.8–2.7	40	Up to 12	0	
3	2.8–3.7	Up to 55	Up to 12	0	
4	3.8–4.7	Up to 70	Up to 14	0	
5	4.8–5.7	Up to 80	Up to 14	In grade 5 and above, only 1 and only when needed.	
6	5.8–6.7	Up to 80	Up to 14	1	
7	6.8–7.7	Up to 90	Up to 16	1	
8	7.8–8.7	Up to 90	Up to 16	1	
9	8.8–9.7	Up to 90	Up to 16	1	
10–12	9.8–10.7	Up to 100	Up to 16	1	

Level of Difficulty: Cognitive Load, Content Differentiation, and Presentation

In addition to readability, each item is constructed with consideration to cognitive load, content differentiation, and presentation as appropriate for the ability and experience of a typical student at that grade level.

- **Cognitive Load:** Cognitive load involves the type and amount of knowledge and thinking that a student must have and use in order to answer the item correctly. The entire impact of the stem and answer choices must be considered.
- **Content Differentiation:** Content differentiation involves the level of detail that a student must address to correctly answer the item. Determining and/

or selecting the correct answer should not be dependent on noticing subtle differences in the stem or answer choices.

- **Depth of Knowledge:** Depth of Knowledge is a language system used as an evaluative tool for differentiating among the different levels, 1 through 4, of complexity of specific learning expectations. Items are written to engage students at the targeted depth of knowledge identified for each skill within the assessment.
- **Presentation:** The presentation of the item includes consistent placement of item components, including directions, stimulus components, questions, and answer choices. Each of these should have a typical representation for the discipline area and grade level. The level of visual differentiation needed to read and understand the item components must be grade-level appropriate.

Efficiency in Use of Student Time

Efficiency is evidenced by a good return of information in relation to the amount of time the student spends on the item. The action(s) required of the student are clearly evident. Ideally, the student is able to answer the question without reading the answer choices. Star Reading items have clear, concise, precise, and straightforward wording.

Balanced Items: Bias and Fairness

Item development meets established demographic and contextual goals that are monitored during development to ensure the item bank is demographically and contextually balanced. Goals are established and tracked in the following areas: use of fiction and nonfiction text, subject and topic areas, geographic region, gender, ethnicity, occupation, age, and disability.

- Items are free of stereotyping, representing different groups of people in non-stereotypical settings.
- Items do not refer to inappropriate content that includes but is not limited to content that presents stereotypes based on ethnicity, gender, culture, economic class, or religion.
- Items do not present any ethnicity, gender, culture, economic class, or religion unfavorably.
- Items do not introduce inappropriate information, settings, or situations.
- Items do not reference illegal activities, sinister or depressing subjects, religious activities or holidays based on religious activities, witchcraft, or unsafe activities.

Accuracy of Content

Concepts and information presented in items are accurate, up-to-date, and verifiable. This includes, but is not limited to, references, dates, events, and locations.

Language Conventions

Grammar, usage, mechanics, and spelling conventions in all Star Reading items adhere to the rules and guidelines in the approved content reference books. *Merriam Webster's 11th Edition* is the reference for pronunciation and spelling. *The Chicago Manual of Style 17th Edition* is the anchor reference for grammar, mechanics, and usage.

Item Components

In addition to the guidelines outlined above, there are criteria that apply to individual item components. The guidelines for passages are addressed above. Specific considerations regarding stem and distractors are listed below.

Item stems meet the following criteria with limited exceptions:

- The question is concise, direct, and a complete sentence. The question is written so students can answer it without reading the distractors.
- Generally, completion (blank) stems are not used. If a completion stem is necessary, (such as is the case with vocabulary in context skills) the stem contains enough information for the student to complete the stem without reading the distractors, and the completion blank is as close to the end of the stem as possible.
- The stem does not include verbal or other clues that hint at correct or incorrect distractors.
- The syntax and grammar are straightforward and appropriate for the grade level. Negative construction is avoided.
- The stem does not contain more than one question or part.
- Concepts and information presented in the items are accurate, up-to-date, and verifiable. This includes but is not limited to dates, references, locations, and events.

Distractors meet the following criteria with limited exceptions:

- All distractors are plausible and reasonable.

- Distractors do not contain clues that hint at correct or incorrect distractors. Incorrect answers are created based on common student mistakes.
- Distractors that are not common mistakes may vary between being close to the correct answer or close to a distractor that is the result of a common mistake.
- Distractors are independent of each other, are approximately the same length, have grammatically parallel structure, and are grammatically consistent with the stem.
- *None of these, none of the above, not given, all of the above,* and *all of these* are not used as distractors.

Metadata Requirements and Goals

Due to the restrictions for modifying text, the content may not meet the following goals; however, new item development works to bring the content into alignment with these goals:

- **Gender**: After removing gender-neutral items, an equal number of male and female items should be represented. In addition to names (Sara) and nouns (sisters), gender is also represented by pronoun (she). Gender is not indicated by subject matter or appeal. For instance, an item on cooking is not female unless there is a female character in it.
- **Ethnicity**: The goal is to provide students with an assessment that reflects the ethnic diversity of our school children within the US: 48% White, 15% Black or African American, 27% Hispanic, 5% Middle Eastern, and 5% Asian or Indian. Ethnicity can be based on name or subject matter.
- **Subject**: A variety of subject areas should be present across the items, such as Arts/Humanities, Science, History, Physical Education, Math, and Technology.

Metadata is tagged with codes for Genres, Ethnicity, Occupations, Subjects, Topics, and Regions.

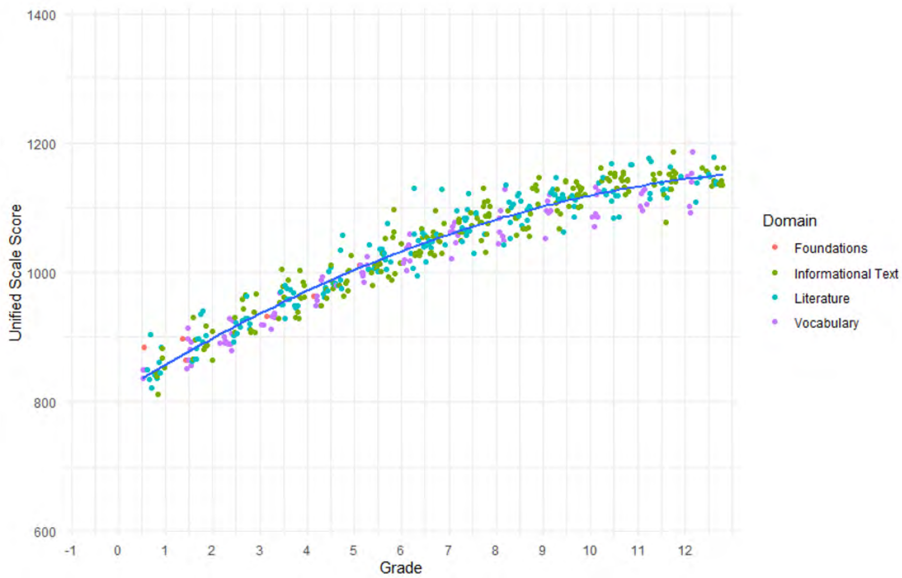
Star Reading and Renaissance Learning Progressions for Reading

Star Reading bridges assessment and instruction through research-based learning progressions to help teachers make effective instructional decisions and to adjust instruction to meet the needs of student at different achievement levels. Star Reading assesses more than 600 grade-specific blueprint skills with items developed and aligned to each skill. The skills measured by Star Reading are drawn from an overarching pool of skills known as the universal skills pool. The universal skills pool contains the full range of skills reflected in state content standards from all 50 US states and the District of Columbia from early literacy to high-school level analysis and critique. The universal skills pool continues to grow and evolve as state standards change and are updated. Learning progressions are created by mapping the skills in the universal skills pool to different content standards. Learning progressions define coherent and continuous pathways in which students acquire knowledge and skills and present the knowledge and skills in teachable orders that can be used to inform instructional decisions.

The first learning progression created for Star Reading was the Renaissance Core Progress for Reading Learning Progression, which identifies a continuum of reading skills that span from early literacy through high-school level analysis and critique. It was developed in consultation with leading experts in early literacy and reading by reviewing research and curricular documents and standards, including the National Assessment of Education Progress (NAEP) Reading framework, Texas Essential Knowledge and Skills, and state reading standards. The Renaissance Core Progress for Reading Learning Progression is supported by calibration data and psychometric analyses and is regularly refined and updated. Item calibration data from Star Reading continually shows that there is a strong correlation between rank ordering of skills in the Renaissance Core Progress for Reading Learning Progression and the item difficulty estimates of items written to measure those skills that are used in Star Reading.

Figure 1 illustrates the relationship between the sequential order of skills in the Renaissance Core Progress for Reading Learning Progression and the average difficulty of the Star Early Literacy and Star Reading items measuring that skill on the Star Reading Unified scale. Each skill is represented by a single data point with skills in each learning progression domain represented by different color points. The figure shows that skills that are ordered later in the Renaissance Core Progression for Reading Learning Progression are often more difficult than skills that are represented earlier in the progression.

Figure 1: Renaissance Core Progress for Reading Learning Progression



The relationships shown in Figure 1continue to evolve as the validation process is ongoing and new items continue to be written. The continual updating of the Renaissance Core Progress for Reading Learning Progression is important to ensure that the ordering of the skills in the Renaissance Core Progress for Reading Learning Progression is an accurate representation of the order in which students learn early literacy and reading skills and concepts. To that end, item calibration data collected from Star Reading is continuously used to validate and refine learning progressions.

Renaissance now develops individualized learning progressions for all 50 states and the District of Columbia. These state specific learning progressions are also updated as state standards change. The state-specific learning progressions cover specific skills represented in each state's grade-level content standards. To create these state-specific learning progressions, each state's content standards are analyzed, tagged, and mapped to skills in the universal skills pool. When standards address areas of learning not yet addressed in the universal skills pool, new skills are developed and added to the universal skills pool and potentially added as new Star Reading skills. Since Star Reading CAT items are written to specific skills which are in turn mapped to skills in the universal skills pool, this allows data from Star Reading CAT items to inform state specific learning progression and allows Star Reading to report results on state specific content standards and learning progressions. This mapping of Star Reading CAT items to skills in the universal skills pool which are in turn mapped to each state's grade-level content standards is one way in which Renaissance works to ensure alignment between Star Reading and state content standards.

When a student completes a Star Reading assessment, the program uses that student's performance to place the student at the appropriate point in the learning progression designated for that school. This learning progression is usually the state specific learning progression for the state in which the school is located. Locating students in the learning progression helps teachers to identify the skills that students are likely to have already learned and the skills they are ready to learn next. It also indicates whether students are meeting the grade-level performance expectations established by state content standards.

Item and Scale Calibration

Background

Star Reading was initially published in 1996, and quickly became one of the first applications of computerized adaptive testing (CAT) to educational assessment at the primary and secondary school levels. Unlike other early CAT applications, the initial version of Star Reading was not based on item response theory (IRT). Instead, it was an instance of stratified adaptive testing (Weiss, 1973[1]). The items in its item bank were sorted into grade levels (strata) based on their vocabulary levels. Examinees started the test at the stratum corresponding to their school grade; an algorithm branched them to easier or more difficult levels, contingent on their performance.

IRT was introduced in Version 2 of Star Reading. At that time, hundreds of new test items were developed, and both the new and the original items from Version 1 were calibrated as to difficulty on a vertical scale using the Rasch model. Star Reading uses the calibrated Rasch difficulty of the test items as the basis for adaptive item selection. And it uses the Rasch difficulty of the items administered to a student, along with the pattern of right and wrong answers, to calculate a maximum likelihood estimate of the location of the student on the Rasch scale. To provide continuity with the non-IRT score scale of Version 1, equipercentile equating was used to transform the Rasch scores to the original Star Reading score scale.

Version 2's Rasch model-based scale of item difficulty and student ability has continued in use in all subsequent versions of Star Reading. This chapter begins by presenting technical details of the development of that Rasch scale. Later, it will describe improvements that have been made to the method of calibrating the Rasch difficulty of new items. Finally, it will present details of the development of a new scale for reporting Star Reading test scores—the Unified Score Scale, first introduced in the 2017–2018 school year.

Calibration of Star Reading Items for Use in Version 2

This section summarizes the psychometric research and development undertaken to prepare the large pool of calibrated reading test questions first used in Star

1. Weiss, D.J. *The stratified adaptive computerized ability test* (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Method Program, 1973. https://eric.ed.gov/?id=ED084301

Reading 2, as well as the linkage of Star Reading 2 scores to the original Star Reading 1 score scale. This research took place in two stages: item calibration and score scale calibration. These are described in their respective sections below.

In Star Reading 2 development, a large-scale item calibration program was conducted in the spring of 1998. The Star Reading 2 item calibration study incorporated all of the newly written vocabulary-in-context and authentic text passage items, as well as over 800 vocabulary items in the Star Reading 1 item bank. Two distinct phases comprised the item calibration study. The first phase was the collection of item response data from a multi-level national student sample. The second phase involved the fitting of item response models to the data, and developing a single IRT difficulty scale spanning all levels from grades 1–12.

Sample Description

The data collection phase of the Star Reading 2 calibration study began with a total item pool of over 2100 items. A nationally representative sample of students tested these items. A total of 27,807 students from 247 schools participated in the item calibration study. Table 9 provides the numbers of students in each grade who participated in the study.

Table 9: Numbers of Students Tested by Grade, Star Reading 2 Item Calibration Study—Spring 1998

Grade Level	Number of Students Tested	Grade Level	Number of Students Tested	Grade Level	Number of Students Tested
1	4,037	5	2,167	9	2,030
2	3,848	6	1,868	10	1,896
3	3,422	7	1,126	11	1,326
4	3,322	8	713	12	1,715
				Not Given	337

Table 10 presents descriptive statistics concerning the makeup of the calibration sample. This sample included 13,937 males and 13,626 females (244 student records did not include gender information). As Table 10 illustrates, the tryout sample approximated the national school population fairly well.

Table 10: Sample Characteristics, Star Reading 2 Calibration Study—Spring 1998 (N = 27,807 Students)

		Students	
		National %	Sample %
Geographic Region	Northeast	20%	16%
	Midwest	24%	34%
	Southeast	24%	25%
	West	32%	25%
District Socioeconomic Status	Low: 31–100%	30%	28%
	Average: 15–30%	29%	26%
	High: 0–14%	31%	32%
	Non-Public	10%	14%
School Type & District Enrollment	Public < 200 200–499 500–2,000 > 2,000	17% 19% 27% 28%	15% 21% 25% 24%
	Non-Public	10%	14%

Table 11 provides information about the ethnic composition of the calibration sample. As Table 11 shows, the students participating in the calibration sample closely approximate the national school population.

Table 11: Ethnic Group Participation, Star Reading 2 Calibration Study—Spring 1998 (N = 27,807 Students)

		Students	
		National %	Sample %
Ethnic Group	Asian	3%	3%
	Black	15%	13%
	Hispanic	12%	9%
	Native American	1%	1%
	White	59%	63%
	Unclassified	9%	10%

Item Presentation

For the calibration research study, seven levels of test booklets were constructed corresponding to varying grade levels. Because reading ability and vocabulary growth are much more rapid in the lower grades, only one grade was assigned per test level for the first four levels of the test (through grade 4). As grade level increases, there is more variation among both students and school curricula, so a single test can cover more than one grade level. Grades were assigned to test levels after extensive consultation with reading instruction experts as well as considering performance data for items as they functioned in the Star Reading 1 test. Items were assigned to grade levels such that the resulting test forms sampled an appropriate range of reading ability typically represented at or near the targeted grade levels.

Grade levels corresponding to each of the seven test levels are shown in the first two columns of Table 12. Students answered a set number of questions at their current grade level, as well as a number of questions one grade level above and one grade level below their grade level. Anchor items were included to support vertically scaling the test across the seven test levels. Table 12 breaks down the composition of test forms at each test level in terms of types and number of test questions, as well as the number of calibration test forms at each level.

Table 12: Calibration Test Forms Design by Test Level, Star Reading 2 Calibration Study—Spring 1998

Test Level	Grade Levels	Items per Form	Anchor Items per Form	Unique Items per Form	Number of Test Forms
A	1	44	21	23	14
B	2	44	21	23	11
C	3	44	21	23	11
D	4	44	21	23	11
E	5–6	44	21	23	14
F	7–9	44	21	23	14
G	10–12	44	21	23	15

Each of the calibration test forms within a test level consisted of a set of 21 anchor items which were common across all test forms within a test level. Anchor items consisted of items: a) on grade level, b) one grade level above, and c) one grade level below the targeted grade level. The use of anchor items facilitated equating of both test forms and test levels for purposes of data analysis and the development of the overall score scale.

In addition to the anchor items were a set of 23 additional items that were unique to a specific test form (within a level). Items were selected for a specific test level based on Star Reading 1 grade level assignment, EDL vocabulary grade designation, or expert judgment. To avoid problems with positioning effects resulting from the placement of items within each test booklet form, items were shuffled within each test form. This created two variations of each test form such that items appeared in different sequential positions within each “shuffled” test form. Since the final items would be administered as part of a computer-adaptive test, it was important to remove any effects of item positioning from the calibration data so that each item could be administered at any point during the test.

The number of field test forms constructed for each of the seven test levels is shown in the last column of Calibration Test Forms Design by Test Level, Star Reading 2 Calibration Study—Spring 1998 (varying from 11–15 forms per level). Calibration test forms were spiraled within a classroom such that each student received a test form essentially at random. This design ensured that no more than two or three students in any classroom attempted any particular tryout item. Additionally, it ensured a balance of student ability across the various tryout forms. Typically, 250–300 students at the designated grade level of the test item received a given question on their test.

It is important to note that some performance data already existed for the majority of the questions in the Star Reading 2 calibration study. All of the questions from the Star Reading 1 item bank were included, as were many items that were previously field tested, but were not included in the Star Reading 1 test.

Following extensive quality control checks, the Star Reading 2 calibration research item response data were analyzed, by level, using both traditional item analysis techniques and IRT methods. For each test item, the following information was derived using traditional psychometric item analysis techniques:

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

Item Difficulty

The difficulty of an item, in traditional item analysis, is the percentage of students who answer the item correctly. This is typically referred to as the “p-value” of the item. Low p-values (such as 15 percent) indicate that the item is difficult since only a small percentage of students answered it correctly. High p-values (such as 90 percent) indicate that almost all students answered the item correctly, and thus the item is easy. It should be noted that the p-value only has meaning for a particular item relative to the characteristics of the sample of students who responded to it.

Item Discrimination

The traditional measure of the discrimination of an item is the correlation between the “score” on the item (correct or incorrect) and the total test score. Items that correlate well with total test score also tend to correlate well with one another and produce a test that has more reliable scores (more internally consistent). For the correct answer, the higher the correlation between item score and total score, the better the item is at discriminating between low scoring and high scoring students. Such items generally will produce optimal test performance. When the correlation between the correct answer and total test score is low (or negative), it typically indicates that the item is not performing as intended. The correlation between endorsing incorrect answers and total score should generally be low since there should not be a positive relationship between selecting an incorrect answer and scoring higher on the overall test.

Item Response Function

In addition to traditional item analyses, the Star Reading calibration data were analyzed using Item Response Theory (IRT) methods. Although IRT encompasses a family of mathematical models, the Rasch model was selected for the Star Reading 2 data both for its simplicity and its ability to accurately model the performance of the Star Reading 2 items.

IRT attempts to model quantitatively what happens when a student with a specific level of ability attempts to answer a specific question. IRT calibration places the item difficulty and student ability on the same scale; the relationship between them can be represented graphically in the form of an item response function (IRF), which describes the probability of answering an item correctly as a function of the student’s ability and the difficulty of the item.

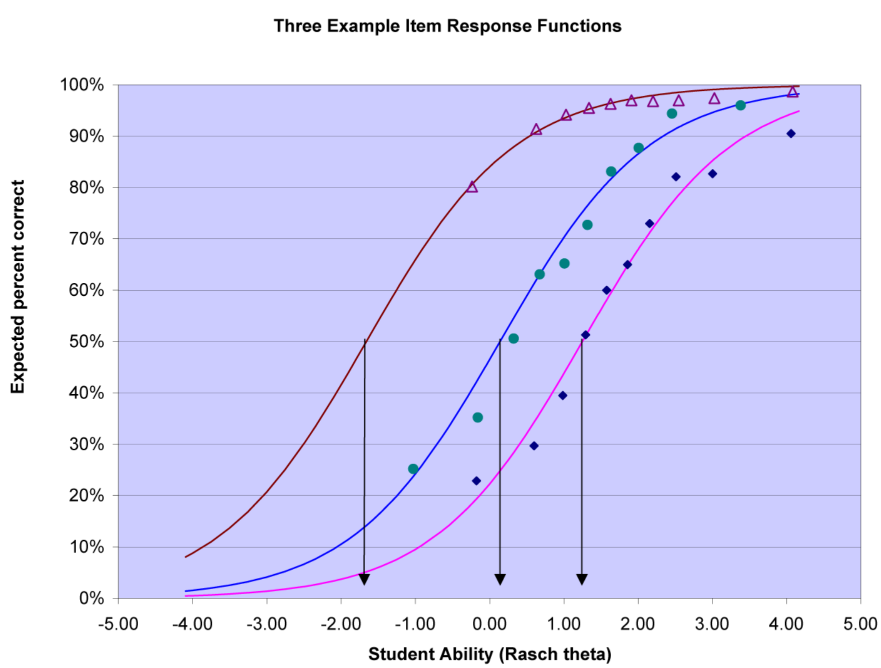
Figure 2 is a plot of three item response functions: one for an easy item, one for a more difficult one, and one for a very difficult item. Each plot is a continuous S-shaped (ogive) curve. The horizontal axis is the scale of student ability, ranging from very low ability (–5.0 on the scale) to very high ability (+5.0 on the scale). The vertical axis is the percent of students expected to answer each of the three items correctly at any given point on the ability scale. Notice that the expected percent correct increases as student ability increases, but varies from one item to another.

In Figure 2, each item's difficulty is the scale point where the expected percent correct is exactly 50. These points are depicted by vertical lines going from the 50 percent point to the corresponding locations on the ability scale. The easiest item has a difficulty scale value of about –1.67; this means that students located at –1.67 on the ability scale have a 50-50 chance of answering that item right. The scale values of the other two items are approximately +0.20 and +1.25, respectively.

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

Also plotted in Figure 2 are "empirical item response functions (EIRF)": the actual percentages of correct responses of groups of students to all three items. Each group is represented as a small triangle, circle, or diamond. Each of those geometric symbols is a plot of the percent correct against the average ability level of the group. Ten groups' data are plotted for each item; the triangular points represent the groups responding to the easiest item. The circles and diamonds, respectively, represent the groups responding to the moderate and to the most difficult item.

Figure 2: Example of Item Statistics Database Presentation of Information



For purposes of the Star Reading 2 calibration research, two different "fit" measures (both unweighted and weighted) were computed. Additionally, if the IRT model is functioning well, then the EIRF points should approximate the (estimated) theoretical IRF. Thus, in addition to the traditional item analysis information, the following IRT-related information was determined for each item administered during the calibration research analyses:

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

Rules for Item Retention

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

Items were eliminated when they met one or more of the following criteria:

- Item-total correlation (item discrimination) was < 0.30
- Some other answer option had an item discrimination that was high
- Sample size of students attempting the item was less than 300
- The traditional item difficulty indicated that the item was too difficult or too easy
- The item did not appear to fit the Rasch model

For Star Reading version 2, after each content reviewer had designated certain items for elimination, their recommendations were combined and a second review was conducted to resolve issues where there was not uniform agreement among all reviewers.

Of the initial 2100+ items administered in the Star Reading 2 calibration research study, 1,409 were deemed of sufficient quality to be retained for further analyses. Traditional item-level analyses were conducted again on the reduced data set that excluded the eliminated items. IRT calibration was also performed on the reduced data set and all test forms and levels were equated based on the information provided by the embedded anchor items within each test form. This resulted in placing the IRT item difficulty parameters for all items onto a single scale spanning grades 1–12.

Table 13 summarizes the final analysis information for the test items included in the calibration test forms by test level (A–G). As shown in the table, the item placements in test forms were appropriate: the average percentage of students correctly answering items is relatively constant across test levels. Note, however, that the average scaled difficulty of the items increases across successive levels of the calibration tests, as does the average scaled ability of the students who answered questions at each test level. The median point-biserial correlation, as shown in the table, indicates that the test items were performing well.

Table 13: Calibration Test Item Summary Information by Test Level, Star Reading 2 Calibration Study—Spring 1998

Test Level	Grade Level(s)	Number of Items	Sample Size	Average Percent Correct	Median Percent Correct	Median Point-Biserial	Average Scaled Difficulty	Average Scaled Ability
A	1	343	4,226	67	75	0.56	–3.61	–2.36
B	2	274	3,911	78	88	0.55	–2.35	–0.07
C	3	274	3,468	76	89	0.51	–1.60	0.76
D	4	274	3,340	69	81	0.51	–0.14	1.53
E	5–6	343	4,046	62	73	0.47	1.02	2.14
F	7–9	343	3,875	68	76	0.48	2.65	4.00
G	10–12	366	4,941	60	60	0.37	4.19	4.72

Scale Calibration and Linking

The outcome of the item calibration study described above was a sizable bank of test items suitable for use in the Star Reading 2 test, with an IRT difficulty scale parameter for each item. The item difficulty scale itself was devised such that it spanned a range of item difficulty from grades 1–12. An important feature of Item Response Theory is that the same scale used to characterize the difficulty of the test items is also used to characterize examinees' ability; in fact, IRT models express the probability of a correct response as a function of the difference between the scale values of an item's difficulty and an examinee's ability. The IRT ability/difficulty scale is continuous; values of observed Rasch ability ranged from about –20 to +20, with the zero value occurring at about the sixth-grade level.

This continuous Rasch score scale is very different from the Scaled Score metric used in Star Reading version 1. Star Reading version 1 scaled scores ranged from 50–1,350, in integer units. The relationship of those scaled scores to the IRT ability scale introduced in Star Reading version 2 was expected to be direct, but not necessarily linear. For continuity between Star Reading 1 and Star Reading 2 scoring, it was desirable to be able to report Star Reading 2 scores on the same scale used in Star Reading 1. To make that possible, a scale linking study was undertaken in conjunction with Star Reading 2 norming. At every grade from 1–12, a portion of the norming sample was asked to take both versions of the Star Reading test: versions 1 and 2. The test score data collected in the course of the linking study were used to link the two scales, providing a conversion table for transforming Star Reading 2 ability scores into equivalent Star Reading 1 Scaled Scores.

From around the country and spanning all 12 grades, 4,589 students participated in the linking study. Linking study participants took both Star Reading 1 and Star Reading 2 tests within a few days of each other. The order in which they took the two test versions was counterbalanced to account for the effects of practice and fatigue. Test score data collected were edited for quality assurance purposes, and 38 cases with anomalous data were eliminated from the linking analyses; the linking was accomplished using data from 4,551 cases. The linking of the two score scales was accomplished by means of an equipercentile equating involving all 4,551 cases, weighted to account for differences in sample sizes across grades. The resulting table of 99 sets of equipercentile equivalent scores was then smoothed using a monotonic spline function, and that function was used to derive a table of Scaled Score equivalents corresponding to the entire range of IRT ability scores observed in the norming study. These Star Reading 2 Scaled Score equivalents range from 0–1400; the same scale has been used for all subsequent Star Reading versions, from version 3 to the present.

Summary statistics of the test scores of the 4,551 cases included in the linking analysis are listed in Table 14. The table lists actual Star Reading 1 Scaled Score means and standard deviations, as well as the same statistics for Star Reading 2 IRT ability estimates and equivalent Scaled Scores calculated using the conversion table from the linking study. Comparing the Star Reading 1 Scaled Score means to the IRT ability score means illustrates how different the two metrics are.

Comparing the Star Reading 1 Scaled Score means to the Star Reading 2 Equivalent Scale Scores in the rightmost two columns of Table 14 illustrates how successful the scale linking was.

Table 14: Summary Statistics of Star Reading 1 and 2 Scores from the Linking Study, by Grade—Spring 1999 (N = 4,551 Students)

Grade Level	Sample Size	Star Reading 1 Scaled Scores		Star Reading 2 IRT Ability Scores		Star Reading 2 Equivalent Scale Scores	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
1	284	216	95	–1.98	1.48	208	109
2	772	339	115	–0.43	1.60	344	148
3	476	419	128	0.33	1.53	419	153
4	554	490	152	0.91	1.51	490	187
5	520	652	176	2.12	1.31	661	213
6	219	785	222	2.98	1.29	823	248
7	702	946	228	3.57	1.18	943	247
8	545	958	285	3.64	1.40	963	276
9	179	967	301	3.51	1.59	942	292
10	81	1,079	292	4.03	1.81	1,047	323
11	156	1,031	310	3.98	1.53	1,024	287
12	63	1,157	299	4.81	1.42	1,169	229
1–12	4,551	656	345	1.73	2.36	658	353

Data from the linking study made it clear that Star Reading 2 software measures ability levels extending beyond the minimum and maximum Star Reading 1 Scaled Scores. In order to retain the superior bandwidth of Star Reading 2 software, extrapolation procedures were used to extend the Scaled Score range below 50 and above 1,350; the range of reported scale scores for Star Reading versions 2 and later is 0 to 1400 for the Enterprise Scale. The Unified Scale reports scores that range from 600 to 1400.

Online Data Collection for New Item Calibration

As described above, beginning with Star Reading Version 2, item calibration involved administering new items and scale anchoring items to national student samples in printed test booklets. Beginning with Star Reading version 4.3, data needed for item calibration have been collected on-line, by embedding small numbers of uncalibrated items within Star Reading tests. After sufficient numbers of item responses have accumulated, the Rasch difficulty of each new item is estimated by fitting a logistic model to the item response data and the Star Reading Rasch scores of the students' tests. Renaissance Learning calls this overall process "dynamic calibration."

Typically, dynamic calibration is done in batches of several hundred new test items. Each student's test may include between 1 and 5 uncalibrated items. Each item is tagged with a grade level, and is typically administered only to students at that grade level and the next higher grade. The selection of the uncalibrated items to be administered to each student is at random, resulting in nearly equivalent distributions of student ability for each item at a given grade level.

Both traditional and IRT item analyses are conducted of the item response data collected. The traditional analyses yielded proportion correct statistics, as well as biserial and point-biserial correlations between scores on the new items and actual scores on the Star Reading tests. The IRT analyses differed from those used in the calibration of Star Reading 2 items, in that the relationships between scores on each new item and the actual Star Reading scores were used to calibrate the Rasch difficulty parameters.

For dynamic calibration, a minimum of 1,000 responses per item is the data collection target. In practice, because of the very large number of Star Reading tests administered each year, the average number of students responding to each new test item is typically several times the target. The calibration analysis proceeds one item at a time, using SAS/STAT™ software to estimate the threshold (difficulty) parameter of every new item by calculating the non-linear regression of each new item score (0 or 1) on the Star Reading Rasch ability estimates. The accuracy of the non-linear regression approach has been corroborated by conducting parallel analyses using Winsteps software. In tests, the two methods yielded virtually identical results.

Table 15 summarizes the final analysis information for the 854 new test items introduced in Star Reading Version 4.3, in 2007, by the target grades tagged to each item. Since that time, several thousand more Star Reading items have gone through dynamic calibration; currently the Star Reading operational item bank contains more than 6,000 items.

Table 15: Calibration Test Item Summary Information by Test Item Grade Level, Star Reading 4.3 Calibration Study–Fall 2007

Item Grade Level	Number of Items	Sample Size[a]	Average Percent Correct	Median Percent Correct	Median Point-Biserial	Average Scaled Difficulty	Average Scaled Ability
K	51	230,580	78	78	47	–3.77	–1.65
1	68	238,578	82	82	45	–3.68	–1.23
2	99	460,175	76	76	51	–2.91	–1.06
3	130	693,184	74	78	47	–1.91	–0.23
4	69	543,554	74	78	41	–1.05	0.64
5	44	514,146	70	72	40	–0.14	1.24
6	32	321,855	71	72	38	0.15	1.62
7	42	402,530	60	58	37	1.40	2.07
8	46	317,110	55	53	33	2.10	2.36
9	36	174,906	54	50	33	2.39	2.59
10	56	99,387	51	54	31	2.95	2.91
11	68	62,596	47	43	22	3.50	3.12
12	51	43,343	44	41	18	3.60	3.11
> 12	62	52,359	34	31	11	4.30	3.10

a. Sample size” in this table is the total number of item responses. Each student was presented with 3, 4, or 5 new items, so the sample size substantially exceeds the number of students.

Computer-Adaptive Test Design

In computer-adaptive tests like the Star Reading test, the items taken by a student are dynamically selected in light of that student’s performance during the testing session. Thus, a low-performing student’s reading skills may branch to easier items in order to better estimate his or her reading achievement level. High-performing students may branch to more challenging reading items in order to better determine the breadth of their reading skills and their reading achievement level.

During a Star Reading test, a student may be “routed” to items at the lowest reading level or to items at higher reading levels within the overall pool of items, depending on the student’s unfolding performance during the testing session. In general, when an item is answered correctly, the student is then given a more difficult item. When an item is answered incorrectly, the student is then given an easier item. Item difficulty here is defined by results of the Star Reading item calibration studies.

Students who have not taken a Star Reading test within six months initially receive an item whose difficulty level is relatively easy for students at the examinee's grade level. The selection of an item that is a bit easier than average minimizes any effects of initial anxiety that students may have when starting the test and serves to better facilitate the student's initial reactions to the test. These starting points vary by grade level and were based on research conducted as part of the national item calibration study.

When a student has taken a Star Reading test within the last 120 days, the difficulty of the first item depends on that student's previous Star Reading test score information. After the administration of the initial item, and after the student has entered an answer, Star Reading software estimates the student's reading ability. The software then selects the next item randomly from among all of the items available that closely match the student's estimated reading ability.

Randomization of items with difficulty values near the student's adjusted reading ability allows the program to avoid overexposure of test items. Items that have been administered to the same student within the past 120 days are not available for administration. The large numbers of items available in the item pools, however, ensure that this constraint has negligible impact on the quality of each Star Reading computer-adaptive test.

Scoring in the Star Reading Tests

Following the administration of each Star Reading item, and after the student has selected an answer, an updated estimate of the student's reading ability is computed based on the student's responses to all items that have been administered up to that point. A proprietary Bayesian-modal Item Response Theory (IRT) estimation method is used for scoring until the student has answered at least one item correctly and one item incorrectly. Once the student has met the 1-correct/1-incorrect criterion, Star Reading software uses a proprietary Maximum-Likelihood IRT estimation procedure for scoring.

This approach to scoring enables Star Reading to provide Scaled Scores that are statistically consistent and efficient. Accompanying each Scaled Score is an associated measure of the degree of uncertainty, called the conditional standard error of measurement (CSEM). The CSEM values for the Star Reading test are unique for each student. CSEM values are dependent on the particular items the student received and on the student's performance on those items.

Scaled Scores are expressed on a common scale that spans all grade levels covered by Star Reading (grades K–12). Because of this common scale, Scaled Scores are directly comparable with each other, regardless of grade level. Other scores, such as Percentile Ranks and Grade Equivalents, are derived from the Scaled Scores.

A New Scale for Reporting Star Reading Test Scores

In 2001, five years following the publication of Star Reading Version 1, Renaissance Learning released Star Early Literacy, an assessment of pre-literacy skills that must be developed in order to learn to read. Although the Early Literacy test measures constructs that are different from those assessed in Star Reading, the two assessments are related developmentally, and scores on the two are moderately highly correlated. Over time, many users of Star Reading have also adopted Star Early Literacy; a frequent practice is to transition children from the Early Literacy assessment to Star Reading when they are ready to take the reading assessment. However, the two assessments had very different score scales, making it difficult to recognize the transition point, and impossible to assess growth in cases where Star Early Literacy was used early in the school year, and replaced by Star Reading later in the same year.

What was needed was a common scale that can be used to report scores on both tests. Such a scale, the Unified Score Scale, has been developed, and was introduced into use in the 2017–2018 school year as an optional alternative scale for reporting achievement on both tests. The Unified Scale is the default scale for reporting test results starting in the 2022–2023 school year.

The Unified Score Scale is derived from the Star Reading Rasch scale of ability and difficulty, which was first introduced with the development of Star Reading Version 2.

The unified Star Early Learning scale was developed by performing the following steps:

- The Rasch scale used by Star Early Literacy was linked (transformed) to the Star Reading Rasch scale.
- A linear transformation of the transformed Rasch scale was developed that spans the entire range of knowledge and skills measured by both Star Early Literacy and Star Reading.

Details of these two steps are presented below.

1. The Rasch scale used by Star Early Literacy was linked to the Star Reading Rasch scale.

   In this step, a linear transformation of the Star Early Literacy Rasch scale to the Rasch scale used by Star Reading was developed, using a method for linear equating of IRT (item response theory) scales described by Kolen and Brennan (2004, pages 162–165).

2. Because Rasch scores are expressed as decimal fractions, and may be either negative or positive, a more user-friendly scale score was developed that uses positive integer numbers only. A linear transformation of the extended Star Reading Rasch scale was developed that spans the entire range of knowledge and skills measured by both Star Early Literacy and Star Reading. The transformation formula is as follows:

   Unified Scale Score = INT (42.93 * Star Reading Rasch Score + 958.74)

   where the Star Reading Rasch score has been extended downwards to values as low as –20.00.

   Following are some features and considerations in the development of that scale, called here the “unified scale.”

   a. The Unified Scale’s range is from 0 to approximately 1400. Anchor points were chosen such that the 0 point is lower than the Star Reading Rasch scale equivalent of the lowest obtainable SEL scale score, and the lowest obtainable Star Early Literacy (SEL) and Star Reading (SR) scale scores correspond to cardinal numbers on the new scale.

      i. The minimum SEL scale score of 300 was set equal to 200 on the Unified Scale.

      ii. An SR scale score of 0 was set equal to 600 on the Unified Scale.

   b. The scale uses integer scale scores. New scale scores from 200 to 1400 correspond respectively to the lowest current SEL scale score of 300, and a point slightly higher than the highest current SR scale score of 1400.

   c. The scale is extensible upwards and downwards. Currently, the highest point on the unified scale is just under 1400; but there is no theoretical limit: If SR content were extended beyond the high school reading level, the range of the new scale can be extended upward without limit, as needed. The lowest point is now set at 200—equivalent to the lowest current SEL scale score (300); but the scale can readily be extended downward as low as 0, if a reason arises to do so.

Further details of the transformation of SEL Rasch scores to the SR Rasch scale may be found in the 2018 edition of the *Star Early Literacy Technical Manual.*

Table 16 contains a table of selected Star Reading Rasch ability scores and their equivalents on the Star Reading and Unified Score scales.

Table 16: Some Star Reading Rasch Scores and Their Equivalents on the Star Reading and Unified Score Scales

Minimum Rasch Score	Star Reading Scaled Score	Unified Scale Score
–8.3500	0	600
—6.2845	50	688
–3.1790	100	822
—2.5030	150	851
–1.9030	200	877
–1.2955	250	903
–0.7075	300	928
–0.1805	350	950
0.3390	400	973
0.7600	450	991
1.2450	500	1012
1.6205	550	1028
1.9990	600	1044
2.3240	650	1058
2.5985	700	1070
2.8160	750	1079
3.0090	800	1087
3.2120	850	1096
3.4570	900	1107
3.7435	950	1119
3.9560	1000	1128
4.0780	1050	1133
4.2120	1100	1139
4.3650	1150	1146
4.5790	1200	1155
4.8280	1250	1166
5.0940	1300	1177

Reliability and Measurement Precision

Measurement is subject to error. A measurement that is subject to a great deal of error is said to be *imprecise*; a measurement that is subject to relatively little error is said to be *reliable*. In psychometrics, the term *reliability* refers to the degree of measurement precision, expressed as a proportion. A test with perfect score precision would have a reliability coefficient equal to 1, meaning that 100 percent of the variation among persons' scores is attributable to variation in the attribute the test measures, and none of the variation is attributable to error. Perfect reliability is probably unattainable in educational measurement; for example, a test with a reliability coefficient of 0.90 is more likely. On such a test, 90 percent of the variation among students' scores is attributable to the attribute being measured, and 10 percent is attributable to errors of measurement. Another way to think of score reliability is as a measure of the consistency of test scores. Two kinds of consistency are of concern when evaluating a test's measurement precision: internal consistency and consistency between different measurements. First, internal consistency refers to the degree of confidence one can have in the precision of scores from a single measurement. If the test's internal consistency is 95 percent, just 5 percent of the variation of test scores is attributable to measurement error.

Second, reliability as a measure of consistency between two different measurements indicates the extent to which a test yields consistent results from one administration to another and from one test form to another. Tests must yield somewhat consistent results in order to be useful; the reliability coefficient is obtained by calculating the coefficient of correlation between students' scores on two different occasions, or on two alternate versions of the test given at the same occasion. Because the amount of the attribute being measured may change over time, and the content of tests may differ from one version to another, the internal consistency reliability coefficient is generally higher than the correlation between scores obtained on different administrations.

There are a variety of methods of estimating the reliability coefficient of a test. Methods such as Cronbach's alpha and split-half reliability are single administration methods and assess internal consistency. Coefficients of correlation calculated between scores on alternate forms, or on similar tests administered two or more times on different occasions, are used to assess alternate forms reliability, or test-retest reliability (stability).

In a computerized adaptive test such as Star Reading, content varies from one administration to another, and it also varies with each student's performance. Another feature of computerized adaptive tests based on Item Response Theory

(IRT) is that the degree of measurement error can be expressed for each student's test individually.

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

Another part of evaluating reliability is looking at the reliability of classification decisions. In many applications of Star Reading, three normative benchmarks, set at the 10th, 25th, and 40th percentile ranks, are used to classify students into the performance categories of intensive intervention, intervention, on watch, and at/above benchmark. These classifications are often used in a response-to-intervention (RTI) and multi-tiered system of supports (MTSS) framework by schools. To show reliability of classifications based on benchmarks, decision accuracy and decision consistency indices can be computed. Like reliability coefficients based on test scores, decision accuracy and consistency indices range from 0 to 1 with values close to 1 indicating more accurate and consistent classifications.

This chapter presents three different types of reliability coefficients: generic reliability, split-half reliability, and alternate forms (test-retest) reliability. This is followed by statistics on the conditional standard error of measurement of Star Reading test scores. The chapter also presents indices of decision accuracy and consistency.

The reliability and measurement error presentation is divided into two sections below: First is a section describing the reliability coefficients, standard errors of measurement, and decision accuracy and consistency indices for the 34-item Star Reading tests. Second, another brief section presents reliability coefficients, standard errors of measurement, and decision accuracy and consistency indices for the 25-item Star Reading progress monitoring tests..

34-Item Star Reading Tests

Generic Reliability

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

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

where σ^2_{error} is the variance of the errors of measurement and σ^2_{total} is the variance of test scores. In Star Reading, the variance of the test scores is easily calculated from Scaled Score data. The variance of the errors of measurement may be estimated from the conditional standard error of measurement (CSEM) statistics that accompany each of the IRT-based test scores, including the Scaled Scores, as depicted below.

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

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

Using this technique with the Star Reading 2018–2019 school year data resulted in the generic reliability estimates shown in Table 17 and Table 18 on page 53. Because this method is not susceptible to error variance introduced by repeated testing, multiple occasions, and alternate forms, the resulting estimates of reliability are generally higher than the more conservative alternate forms reliability coefficients. These generic reliability coefficients are, therefore, plausible upper-bound estimates of the internal consistency reliability of the Star Reading computer-adaptive test.

Generic reliability estimates for scores on the Unified score scale are shown in Table 17; Table 18 lists the reliability estimates for the older Star Reading “Enterprise” scale scores. Results in Table 17 indicate that the overall reliability of the Unified scale scores was about 0.98. Coefficients ranged from a low of 0.94 in grade 5 to a high of 0.97 in grade K. Results based on the Enterprise Scale in Table 18 are slightly lower: the overall reliability of those scale scores was about

0.97; within-grade coefficients ranged from a low of 0.93 in grades 3 to 7 to a high of 0.95 in grades K, 1, 11, and 12.

As both tables show, Star Reading reliability is quite high, grade by grade and overall. Star Reading also demonstrates high test-retest consistency as shown in the rightmost columns of the same tables. Star Reading's technical quality for an interim assessment is on a virtually equal footing with the highest-quality summative assessments in use today.

Split-Half Reliability

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

A split-half reliability coefficient is calculated in three steps. First, the test is divided into two halves, and scores are calculated for each half. Second, the correlation between the two resulting sets of scores is calculated; this correlation is an estimate of the reliability of a half-length test. Third, the resulting reliability value is adjusted, using the Spearman-Brown formula, to estimate the reliability of the full-length test.

In internal simulation studies, the split-half method provided accurate estimates of the internal consistency reliability of adaptive tests, and so it has been used to provide estimates of Star Reading reliability. These split-half reliability coefficients are independent of the generic reliability approach discussed earlier and more firmly grounded in the item response data. Split-half scores were based on all of the 34 items of the Star Reading tests; scores based on the odd- and the even-numbered items were calculated separately. The correlations between the two sets of scores were corrected to a length of 34 items, yielding the split-half reliability estimates displayed in Table 17 and Table 18 on page 53.

Results indicated that the overall split-half reliability of the Unified scores was 0.98. The coefficients ranged from a low of 0.94 in grades 4 to 8 to a high of 0.96 in grade 1. On the Enterprise Scale, the overall split-half reliability of the Enterprise scores was 0.97. The coefficients ranged from a low of 0.92 in grades 4 and 5 to a high of 0.95 in grades K, 1, and 12. These reliability estimates are quite consistent across grades 1-12, and quite high, again a result of the measurement efficiency inherent in the adaptive nature of the Star Reading test.

Alternate Form Reliability

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

Content sampling, temporal changes in individuals' performance, and growth or decline over time can affect alternate forms reliability coefficients, usually making them appreciably lower than internal consistency reliability coefficients.

The alternate form reliability study provided estimates of Star Reading reliability using a variation of the test-retest method. In the traditional approach to test-retest reliability, students take the same test twice, with a short time interval, usually a few days, between administrations. In contrast, the Star Reading alternate form reliability study administered two different tests by avoiding during the second test the use of any items the student had encountered in the first test. All other aspects of the two tests were identical. The correlation coefficient between the scores on the two tests was taken as the reliability estimate.

The alternate form reliability estimates for the Star Reading test were calculated using both the Star Reading Unified scaled scores and the Enterprise scaled scores. Checks were made for valid test data on both test administrations and to remove cases of apparent motivational discrepancies.

Table 17 and Table 18 include overall and within-grade alternate reliability, along with an indication of the average number of days between testing occasions. The average number of days between testing occasions ranged from 91–130 days.

Results indicated that the overall reliability of the scores on the Unified scale was about 0.93. The alternate form coefficients ranged from a low of 0.73 in grade K to a high of 0.87 in grade 9. Results for the Enterprise scale were similar to those of the Unified Scale with an overall reliability of 0.93; its alternate form coefficients ranged from a low of 0.76 in grade K to a high of 0.88 in grades 8, 9, and 10.

Because errors of measurement due to content sampling and temporal changes in individuals' performance can affect this correlation coefficient, this type of reliability estimate provides a conservative estimate of the reliability of a single Star Reading administration. In other words, the actual Star Reading reliability is likely higher than the alternate form reliability estimates indicate.

Table 17: Reliability Estimates from the Star Reading 2018–2019 Data on the Unified Scale

		Reliability Estimates—Unified Scale					
		Generic	Split-Half		Alternate Forms		
Grade	N	ρxx	N	ρxx	N	ρxx	Average Days between Testing
K	50,000	0.97	20,000	0.95	7,000	0.73	91
1	1,000,000	0.96	20,000	0.96	200,00	0.76	100
2	1,000,000	0.96	20,000	0.95	200,000	0.83	114
3	1,000,000	0.95	20,000	0.95	200,000	0.85	113
4	1,000,000	0.95	20,000	0.94	200,000	0.86	115
5	1,000,000	0.94	20,000	0.94	200,000	0.86	115
6	1,000,000	0.95	20,000	0.94	200,000	0.86	117
7	1,000,000	0.95	20,000	0.94	200,000	0.86	121
8	1,000,000	0.95	20,000	0.94	200,000	0.86	120
9	500,000	0.96	20,000	0.95	100,000	0.87	127
10	500,000	0.96	20,000	0.95	100,000	0.86	125
11	200,000	0.96	20,000	0.95	40,000	0.85	130
12	200,000	0.96	20,000	0.95	40,000	0.85	122
Overall	9,450,000	0.98	260,000	0.98	1,887,000	0.93	116

Table 18: Reliability Estimates from the Star Reading 2018–2019 Data on the Enterprise Scale

Grade	N	Reliability Estimates—Enterprise Scale					
		Generic	Split-Half		Alternate Forms		
		ρxx	N	ρxx	N	ρxx	Average Days between Testing
K	50,000	0.95	20,000	0.95	7,000	0.76	91
1	1,000,000	0.95	20,000	0.95	200,000	0.80	100
2	1,000,000	0.94	20,000	0.95	200,000	0.85	114
3	1,000,000	0.93	20,000	0.94	200,000	0.86	113
4	1,000,000	0.93	20,000	0.93	200,000	0.86	115
5	1,000,000	0.93	20,000	0.92	200,000	0.86	115
6	1,000,000	0.93	20,000	0.92	200,000	0.87	117
7	1,000,000	0.93	20,000	0.93	200,000	0.87	121
8	1,000,000	0.94	20,000	0.93	200,000	0.88	120
9	500,000	0.94	20,000	0.94	100,000	0.88	127
10	500,000	0.94	20,000	0.94	100,000	0.88	125
11	200,000	0.95	20,000	0.94	40,000	0.87	130
12	200,000	0.95	20,000	0.95	40,000	0.87	122
Overall	9,450,000	0.97	260,000	0.97	1,887,000	0.93	116

Star Reading was designed to be a standards-based assessment, meaning that its item bank measures skills identified by exhaustive analysis of national and state standards in Reading, from grades K–12. The 34-item Star Reading content covers many more skills than Star Reading versions 1 through 4.3, which administered only 25 items.

The increased length of the current version of Star Reading, combined with its increased breadth of skills coverage and enhanced technical quality, was expected to result in improved measurement precision; this showed up as slightly increased reliability, in both internal consistency reliability and alternate form reliability as shown in the tables above. For comparison, see Table 22 on page 60 and Table 23 on page 61.

Standard Error of Measurement

When interpreting the results of any test instrument, it is important to remember that the scores represent estimates of a student's true ability level. Test scores are not absolute or exact measures of performance. Nor is a single test score infallible in the information that it provides. The standard error of measurement can be thought of as a measure of how precise a given score is. The standard error of measurement describes the extent to which scores would be expected to fluctuate because of chance. If measurement errors follow a normal distribution, an SEM of 17 means that if a student were tested repeatedly, his or her scores would fluctuate within 17 points of his or her first score about 68 percent of the time, and within 34 points (twice the SEM) roughly 95 percent of the time. Since reliability can also be regarded as a measure of precision, there is a direct relationship between the reliability of a test and the standard error of measurement for the scores it produces.

The Star Reading tests differ from traditional tests in at least two respects with regard to the standard error of measurement. First, Star Reading software computes the SEM for each individual student based on his or her performance, unlike most traditional tests that report the same SEM value for every examinee. Each administration of Star Reading yields a unique “conditional” SEM (CSEM) that reflects the amount of information estimated to be in the specific combination of items that a student received in his or her individual test. Second, because the Star Reading test is adaptive, the CSEM will tend to be lower than that of a conventional test, particularly at the highest and lowest score levels, where conventional tests’ measurement precision is weakest. Because the adaptive testing process attempts to provide equally precise measurement, regardless of the student’s ability level, the average CSEMs for the IRT ability estimates are very similar for all students.

Table 19 and Table 20 contain two different sets of estimates of Star Reading measurement error: conditional standard error of measurement (CSEM) and global standard error of measurement (SEM). Conditional SEM was just described; the estimates of CSEM in Table 19 and Table 20 are the average CSEM values observed for each grade.

Global standard error of measurement is based on the traditional SEM estimation method, using internal consistency reliability and the variance of the test scores to estimate the SEM:

$$SEM = SQRT(1 - \rho)\ \sigma_x$$

where

SQRT() is the square root operator

ρ is the estimated internal consistency reliability

σ_x is the standard deviation of the observed scores (in this case, Scaled Scores)

Table 19 and Table 20 summarize the distribution of CSEM values for the 2018–2019 data, overall and by grade level. The overall average CSEM on the Unified scale across all grades was 17 scaled score units and ranged from a low of 16 in grades 1–3 to a high of 17 in grades K and 4–12 (Table 19).The average CSEM based on the Unified scale is similar across all grades. The overall average unified scale score global SEM was 18, slightly higher than the average CSEM. Table 20 shows the average CSEM values on the Enterprise Star Reading scale. Although the adaptive testing process attempts to provide equally precise measurement, regardless of the student's ability level, and the average CSEMs for the IRT ability estimates are very similar for all students, the transformation of the Star Reading IRT ability estimates into equivalent Scaled Enterprise Scores is not linear and the resulting SEMs in the Enterprise Scaled Score metric are less similar.

The overall average CSEM on the Enterprise scale across all grades was 54 scaled score units and ranged from a low of 20 in kindergarten to a high of 71 in grade 8. Unlike the Unified scale, the Enterprise Scale CSEM values vary by grade and increased with grade until grade 8. The global SEMs for the Enterprise scale scores were higher at each grade, and overall, than the average CSEMs; the overall average SEM was 56. This is attributable to the nonlinear transformation of the Star Reading IRT ability estimates into equivalent Enterprise Scaled Scores. The Unified scale, in contrast, is based on a linear transformation of the IRT ability estimates; it eliminates the issues of variable and large CSEM values that are an artifact of the Enterprise Scaled Score nonlinear transformation.

Table 19: Standard Error of Measurement for the 2018–2019 Star Reading Data on the Unified Scale

		Standard Error of Measurement—Unified Scale		
		Conditional		
Grade	Sample Size	Average	Standard Deviation	Global
K	50,000	17	2.3	19
1	1,000,000	16	1.3	18
2	1,000,000	16	1.2	17
3	1,000,000	16	1.3	17
4	1,000,000	17	1.3	17
5	1,000,000	17	1.3	17
6	1,000,000	17	1.4	17
7	1,000,000	17	1.5	17
8	1,000,000	17	1.7	17
9	500,000	17	1.9	18
10	500,000	17	2.1	18
11	200,000	17	2.2	18
12	200,000	17	2.5	18
All	9,450,000	17	1.5	18

Table 20: Standard Error of Measurement for the 2018–2019 Star Reading Data on the Enterprise Scale

Grade	Sample Size	Standard Error of Measurement—Enterprise Scale		
		Conditional		Global
		Average	Standard Deviation	
K	50,000	20	16.5	24
1	1,000,000	24	14.0	27
2	1,000,000	33	13.0	36
3	1,000,000	42	15.5	45
4	1,000,000	50	19.6	55
5	1,000,000	58	22.7	63
6	1,000,000	65	24.4	70
7	1,000,000	69	25.4	75
8	1,000,000	71	26.5	77
9	500,000	70	27.5	77
10	500,000	69	28.8	77
11	200,000	69	29.1	76
12	200,000	67	29.6	75
All	9,450,000	54	27.5	56

Decision Accuracy and Decision Consistency

Decision accuracy is generally defined as the degree to which observed examinee classification decisions on a single assessment would agree with true classifications for a given set of cut scores. There are multiple approaches to estimate decision accuracy. Star Reading uses Rudner's index (Rudner, 2001; 2005) based on item response theory (IRT), which assumes that the maximum likelihood estimate of ability converges to a normal distribution with mean equal to θ and standard deviation equal to the conditional standard error of measurement (CSEM). Mathematically, this index can be computed as:

$$\hat{\tau} = \sum (\hat{\mathbf{P}} * \mathbf{W})/N_e,$$

where Σ denotes the summation of all matrix elements, * denotes element-wise matrix multiplication, N_e is the number of examinees, $\hat{\mathbf{P}}$ is a $N_e \times C$ matrix of expected probabilities with C being the number of performance categories on the assessment, and **W** is a $N_e \times C$ matrix of binary weights used to indicate the observed performance categories on the assessment. The $\hat{\mathbf{P}}$ matrix is defined as:

$$\hat{\mathbf{P}} = \begin{bmatrix} \hat{p}_{11} & \hat{p}_{12} & \cdots & \hat{p}_{1C} \\ \hat{p}_{21} & \hat{p}_{22} & \cdots & \hat{p}_{2C} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{p}_{N_e1} & \hat{p}_{N_e2} & \cdots & \hat{p}_{N_eC} \end{bmatrix}$$

with the expected probability $\hat{p}_{ic}$ in the above matrix estimated as:

$$\hat{p}_{ic} = \phi(\kappa_{ic}, \kappa_{i(c+1)}, \hat{\theta}_i, \hat{\sigma}_{\theta_i}),$$

where $\phi(a, b, \mu, \sigma)$ is the area from a to b under a normal curve with a mean of μ and a standard deviation of σ, $\hat{\theta}_i$ is examinee i's IRT ability estimate, $\hat{\sigma}_{\theta_i}$ is the corresponding CSEM for the ability estimate $\hat{\theta}_i$, and κ_{ic} and $\kappa_{i(c+1)}$ are cut scores with $\kappa_{i1} = -\infty$, κ_{i2} being the cut score separating performance categories 1 and 2, κ_{i3} being the cut score separating performance categories 2 and 3, and so on with the last cut score $\kappa_{i(c+1)} = \infty$. The **W** matrix of weights is defined as:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1C} \\ w_{21} & w_{22} & \cdots & w_{2C} \\ \vdots & \vdots & \cdots & \vdots \\ w_{N_e1} & w_{N_e2} & \cdots & w_{N_eC} \end{bmatrix},$$

where the weight, w_{ic}, equals 1 if the student was classified into performance level category C based on their ability estimate and 0 otherwise.

A counterpart to decision accuracy is decision consistency, defined as the degree to which examinees would be classified into the same performance categories given parallel replications of the same assessment. The method used to estimate decision consistency is based on an extension to Rudner's decision accuracy index, which is described in Wyse and Hao (2012). This index can be estimated as:

$$\hat{\gamma} = \sum (\hat{\mathbf{P}} * \hat{\mathbf{P}}) / N_e,$$

where N_e is the number of examinees and $\hat{\mathbf{P}}$ is the same $N_e \times C$ matrix of expected probabilities used when computing the decision accuracy index.

For Star Reading, three different classification decisions based on benchmarks set at the 10th, 25th, and 40th percentile ranks in the student norms are available by default in the Star Reading software. These cut scores are used to separate students into four different performance categories: intensive intervention, intervention, on watch, and at/above benchmark. Table 21 shows estimates of decision accuracy and consistency when identifying students based on the three individual benchmarks as well as all three benchmarks together using random samples of students that took Star Reading in the 2018–2019 school year.

Results indicate that decision accuracy and consistency were quite high overall and across grades. For PR10, decision accuracy ranged from a low of 0.95 to a high of 0.99, while decision consistency ranged from 0.93 to 0.99. For PR25, decision accuracy ranged from a low of 0.93 to a high of 0.97, while decision consistency ranged from 0.90 to 0.96. For PR40, decision accuracy ranged from a low of 0.92 to a high of 0.95, while decision consistency ranged from 0.89 to 0.93. Decision accuracy when using all three benchmarks together ranged from a low of 0.81 to a high of 0.93, while decision consistency ranged from a low of 0.74 to a high of 0.89. These are high levels of decision accuracy and consistency when making classification decisions based on each individual benchmark or all three benchmarks together, and support using Star Reading in RTI/MTSS frameworks.

Table 21: Decision Accuracy and Consistency for Different Benchmarks Based on 2018–2019 Star Reading Tests

Grade	N	Decision Accuracy				Decision Consistency			
		PR10	PR25	PR40	All 3 Benchmarks	PR10	PR25	PR40	All 3 Benchmarks
K	50,000	0.99	0.97	0.95	0.92	0.99	0.96	0.93	0.89
1	1,000,000	0.99	0.96	0.94	0.89	0.98	0.94	0.92	0.85
2	1,000,000	0.97	0.95	0.94	0.86	0.96	0.93	0.91	0.81
3	1,000,000	0.97	0.94	0.94	0.84	0.94	0.92	0.90	0.79
4	1,000,000	0.97	0.94	0.92	0.83	0.95	0.92	0.89	0.77
5	1,000,000	0.96	0.93	0.92	0.82	0.95	0.90	0.89	0.75
6	1,000,000	0.95	0.93	0.92	0.81	0.94	0.90	0.89	0.74
7	1,000,000	0.96	0.93	0.92	0.81	0.94	0.90	0.89	0.75
8	1,000,000	0.96	0.93	0.92	0.81	0.94	0.90	0.89	0.74
9	500,000	0.95	0.93	0.93	0.81	0.93	0.90	0.90	0.74
10	500,000	0.95	0.93	0.93	0.81	0.93	0.90	0.90	0.74
11	200,000	0.95	0.93	0.93	0.81	0.93	0.90	0.90	0.74
12	200,000	0.95	0.93	0.93	0.81	0.93	0.90	0.91	0.74
Overall	9,450,000	0.96	0.94	0.93	0.83	0.95	0.91	0.90	0.77

25-Item Star Reading Progress Monitoring Tests

Star Reading is used for both universal screening and progress monitoring. The 34-item Star Reading test is widely used for universal screening. A shorter version—the 25-item Star Reading progress monitoring test—exists for use in progress monitoring. The following section summarizes the reliability and the standard error of measurement of the progress monitoring version of Star Reading.

Reliability Coefficients

Table 22 and Table 23 show the reliability estimates of the Star Reading progress monitoring test on both the Unified scale and the Enterprise scale using data from the 2017–2018 and 2018–2019 school years.

Table 22: Reliability Estimates from the 2017–2018 and 2018–2019 Star Reading Progress Monitoring Tests on the Unified Scale

	Progress Monitoring Reliability Estimates—Unified Scale			
	Generic		Split-Half	
Grade	N	ρxx	N	ρxx
1	10000	0.94	9400	0.94
2	30000	0.91	29000	0.92
3	30000	0.89	31000	0.90
4	30000	0.88	29000	0.89
5	28500	0.87	26000	0.88
6	14000	0.88	14099	0.90
7	10000	0.89	9400	0.91
8	10000	0.91	9400	0.93
9	1800	0.90	1619	0.93
10	1450	0.92	1376	0.93
11	730	0.93	686	0.96
12	480	0.94	444	0.96
Overall	166,960	0.96	161424	0.96

Table 23: Reliability Estimates from the 2017–2018 and 2018–2019 Star Reading Progress Monitoring Tests on the Enterprise Scale

Grade	Progress Monitoring Reliability Estimates—Enterprise Scale			
	Generic		Split-Half	
	N	ρxx	N	ρxx
1	10,000	0.94	9,400	0.94
2	30,000	0.92	29,000	0.92
3	30,000	0.90	31,000	0.89
4	30,000	0.89	29,000	0.88
5	28,500	0.88	26,000	0.87
6	14,000	0.89	14,099	0.88
7	10,000	0.91	9,400	0.89
8	10,000	0.93	9,400	0.91
9	1800	0.93	1619	0.91
10	1450	0.94	1376	0.92
11	730	0.95	686	0.94
12	480	0.96	444	0.95
Overall	166,960	0.94	161,424	0.94

The progress monitoring Star Reading reliability estimates are also quite high and consistent across grades 1–12, for a test composed of only 25 items.

Overall, these coefficients also compare very favorably with the reliability estimates provided for other published reading tests, which typically contain far more items than the 25-item Star Reading progress monitoring tests. The Star Reading progress monitoring test's high reliability with minimal testing time is a result of careful test item construction and an effective and efficient adaptive-branching procedure.

Standard Error of Measurement

Table 24 and Table 25 show the conditional standard error of measurement (CSEM) and the global standard error of measurement (SEM), overall and by grade level.

Table 24: Estimates of 2017–2018 and 2018–2019 Star Reading Progress Monitoring Measurement Precision by Grade and Overall, on the Unified Scale

Grade	Progress Monitoring Standard Error of Measurement—Unified Scale				
	Conditional			Global	
	Sample Size	Average	Standard Deviation	Sample Size	SEM
1	10,000	19	2.1	9,400	19
2	30,000	19	1.2	29,000	18
3	30,000	19	1.4	31,000	18
4	30,000	19	1.5	29,000	19
5	28,500	19	1.5	26,000	19
6	14,000	19	1.5	14,099	18
7	10,000	19	1.5	9,400	19
8	10,000	19	1.6	9,400	19
9	1,800	19	1.6	1,619	18
10	1,450	19	1.9	1,376	19
11	730	19	1.7	686	18
12	480	19	2.2	444	18
All	166,960	19	1.5	161,424	18

Table 25: Estimates of 2017–2018 and 2018–2019 Star Reading Progress Monitoring Measurement Precision by Grade and Overall, on the Enterprise Scale

Grade	Progress Monitoring Standard Error of Measurement—Enterprise Scale				
	Conditional			Global	
	Sample Size	Average	Standard Deviation	Sample Size	SEM
1	10,000	21	15.3	9,400	25
2	30,000	34	12.9	29,000	36
3	30,000	44	12.9	31,000	45
4	30,000	52	17.4	29,000	55
5	28,500	59	20.5	26,000	63
6	14,000	64	22.4	14,099	67
7	10,000	71	26.0	9,400	76
8	10,000	78	28.8	9,400	82
9	1,800	78	29.1	1,619	78
10	1,450	78	32.7	1,376	85
11	730	78	32.7	686	79
12	480	77	33.7	444	81
All	166,960	51	24.1	161,424	56

Comparing the estimates of reliability and measurement error of Star Reading (Table 17, Table 18, Table 19, and Table 20) with those of Star Reading progress monitoring (Table 23, Table 24, Table 25, and Table 25) confirms that Star Reading is slightly superior to the shorter Star Reading progress monitoring assessments in terms of reliability and measurement precision.

Decision Accuracy and Consistency

Table 26 shows the decision accuracy and consistency indices for PR10, PR25, and PR40 benchmarks for Star Reading Progress Monitoring based on data collected in the 2017-2018 and 2018-2019 school years. Results suggest that the decision accuracy and consistency for the Star Reading Progress Monitoring tests was high, but slightly lower than the values observed for the 34-item Star Reading tests. These high levels of decision accuracy and consistency support using Star Reading tests in RTI/MTSS frameworks.

Table 26: Decision Accuracy and Consistency for Different Benchmarks Based on 2017–2018 and 2018–2019 Star Reading Progress Monitor Tests

Grade	N	Decision Accuracy				Decision Consistency			
		PR10	PR25	PR40	All 3 Benchmarks	PR10	PR25	PR40	All 3 Benchmarks
1	10,000	0.96	0.92	0.91	0.79	0.95	0.88	0.87	0.73
2	30,000	0.94	0.91	0.92	0.77	0.92	0.87	0.88	0.70
3	30,000	0.94	0.90	0.89	0.75	0.92	0.86	0.85	0.67
4	30,000	0.93	0.89	0.89	0.73	0.91	0.84	0.85	0.64
5	28,500	0.93	0.88	0.90	0.71	0.90	0.83	0.86	0.63
6	14,000	0.90	0.88	0.92	0.71	0.86	0.84	0.88	0.62
7	10,000	0.91	0.90	0.92	0.74	0.87	0.85	0.89	0.65
8	10,000	0.92	0.90	0.93	0.76	0.89	0.87	0.90	0.68
9	1,800	0.91	0.91	0.93	0.76	0.87	0.87	0.91	0.68
10	1,450	0.91	0.92	0.93	0.77	0.88	0.89	0.90	0.70
11	730	0.93	0.92	0.93	0.78	0.90	0.89	0.90	0.71
12	480	0.93	0.92	0.20	0.78	0.91	0.89	0.89	0.72
Overall	166,960	0.93	0.90	0.90	0.74	0.91	0.85	0.87	0.66

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.

Content Validity

One touchstone is content validity, which is the relevance of the test questions to the attributes or dimensions intended to be measured by the test—namely reading comprehension, reading vocabulary, and related reading skills, in the case of the Star Reading assessments. The content of the item bank and the content balancing specifications that govern the administration of each test together form the foundation for "content validity" for the Star Reading assessments. These content validity issues were discussed in detail in "Content and Item Development" and were an integral part of the test items that are the basis of Star Reading today.

Construct Validity

Construct validity, which is the overarching criterion for evaluating a test, investigates the extent to which a test measures the construct(s) that it claims to be assessing. Establishing construct validity involves the use of data and other information external to the test instrument itself. For example, Star Reading claims to provide an estimate of a child's reading comprehension and achievement level. Therefore, demonstration of Star Reading's construct validity rests on the evidence that the test provides such estimates. There are a number of ways to demonstrate this.

For instance, in a study linking Star Reading Version 1 and the Degrees of Reading Power comprehension assessment, a raw correlation of 0.89 was observed between the two tests. Adjusting that correlation for attenuation due to unreliability yielded a corrected correlation of 0.96 between the two assessments, indicating that the constructs measured by the different tests are essentially indistinguishable.

Since reading ability varies significantly within and across grade levels and improves as a student's grade placement increases, scores within Star Reading should demonstrate these anticipated internal relationships; in fact, they do. Additionally, scores for Star Reading should correlate highly with other accepted procedures and measures that are used to determine reading achievement and reading comprehension; this is external construct validity. This section deals with both internal and external evidence of the validity of Star Reading as an assessment of reading comprehension and reading skills.

Internal Evidence: Evaluation of Unidimensionality of Star Reading

Star Reading is a 34-item computerized-adaptive assessment that measures reading comprehension. Its items are selected adaptively for each student, from a very large bank of reading test items, each of which is aligned to one of five blueprint domains:

- Word knowledge and skills,
- Comprehension strategies and constructing meaning,
- Analyzing literary text,
- Analyzing argument and evaluating text, and
- Understanding author's craft.

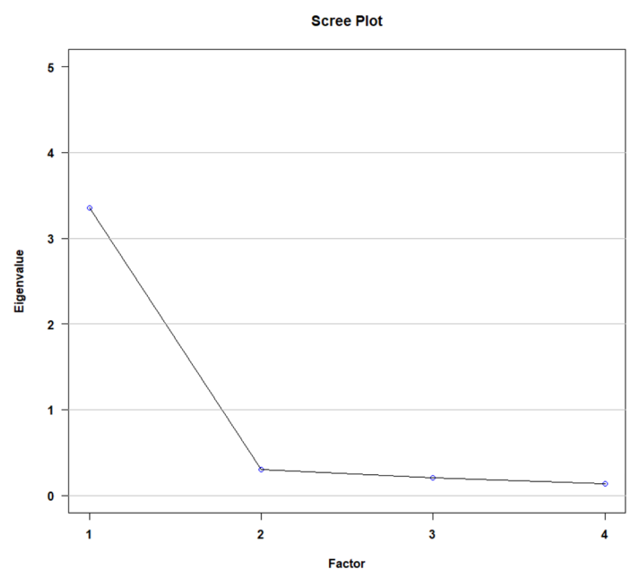
Star Reading is an application of item response theory (IRT); each test item's difficulty has been calibrated using the Rasch model. One of the assumptions of the Rasch model is unidimensionality: that a test measures only a single construct such as reading comprehension in the case of Star Reading. To evaluate whether Star reading measures a single construct, factor analyses were conducted. Factor analysis is a statistical technique used to determine the number of dimensions or constructs that a test measures. Both exploratory and confirmatory factor analyses were conducted across grades K to 12.

To begin, a large sample of student Star Reading data was assembled. The overall sample consisted of 286,000 student records. That sample was divided into 2 sub-samples. The first sub-sample, consisting of 26,000 cases, was used for exploratory factor analysis; the second sub-sample, 260,000 cases, was reserved for confirmatory factor analyses that followed the initial exploratory analysis.

Within each sub-sample, each student's 34 Star Reading item responses were divided into subsets of items aligned to each of the 5 blueprint domains. Tests administered in grades 4–12 included items from all five domains. Tests given in grades K–3 included items from just 4 domains; no items measuring analyzing argument and evaluating text were administered in these grades. For each student, separate Rasch ability estimates (subtest scores) were calculated from each domain-specific subset of item responses. A Bayesian sequential procedure developed by Owen (1969, 1975) was used for the subtest scoring. The number of items included in each subtest ranged from 2 to 18, following the Star Reading test blueprints, which specify different numbers of items per domain, depending on the student's grade level.

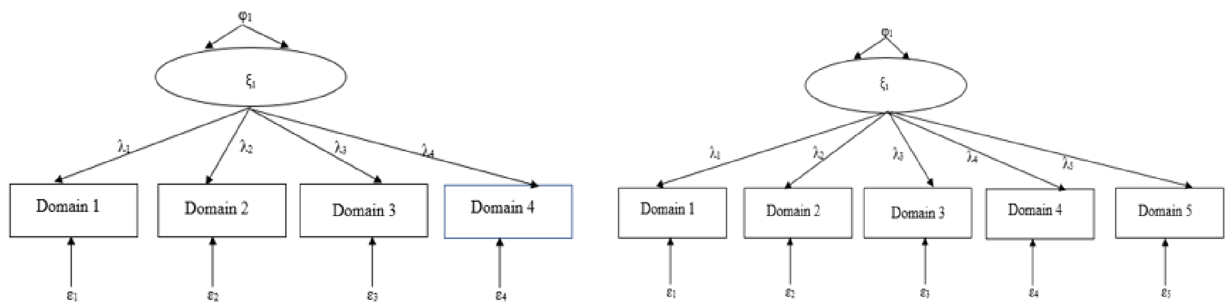
Intercorrelations of the blueprint domain-specific Rasch subtest scores were analyzed using exploratory factor analysis (EFA) to evaluate the number of dimensions/ factors underlying Star Reading. Varimax rotation was used. In each grade, the EFA analyses retained a single dominant underlying dimension based on either the MINEIGEN (eigenvalue greater than 1) or the PROPORTION criterion (proportion of variance explained by the factor), as expected. An example of a scree plot from grade 2 based on the PROPORTION criterion is shown in Figure 3.

Figure 3: Example Scree Plot from the Grade 2 Exploratory Factor Analysis in Star Reading



Subsequent to the EFA analyses, confirmatory factor analyses (CFA) were also conducted using the subtest scores from the CFA sub-sample. A separate confirmatory analysis was conducted for each grade. The CFA models tested a single underlying model as shown in Figure 4. Two CFA models were fitted because one of the Star Reading blueprint domains is not tested in grades K to 3.

Figure 4: Confirmatory Factor Analyses (CFA) in Star Reading



Model 1: Grade K to Grade3 Model 2: Grade 4 to Grade 12

The results of the CFA analyses are summarized in Table 27. As that table indicates, sample sizes ranged from 18,723 to 20,653; because the chi-square (X^2) test is not a reliable test of model fit when sample sizes are large, fit indices are presented. The comparative fit index (CFI) and Tucker-Lewis Index (TLI) are shown; for these indices, values are either 1 or very close to 1, indicating strong evidence of a single construct/dimension. In addition, the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) are presented. RMSEA and SRMR values less than 0.08 indicate good fit. Cutoffs for the indices are presented in Hu and Bentler, 1999. Overall, the CFA results strongly support a single underlying construct in Star Reading.

Table 27: Summary of the Goodness-of-Fit of the CFA Models for Star Reading by Grade

Grade	N	X^2	df	CFI	TLI	RMSEA	SRMR
K	20,000	16.005	2	1.000	1.000	0.019	0.002
1	20,000	8.716	2	1.000	1.000	0.013	0.002
2	20,000	34.23	2	1.000	0.999	0.028	0.003
3	20,000	34.982	2	1.000	0.999	0.029	0.003
4	20,000	109.821	5	0.999	0.997	0.032	0.005
5	20,000	53.772	5	0.999	0.999	0.022	0.004
6	20,000	127.682	5	0.998	0.997	0.035	0.006
7	20,000	154.811	5	0.998	0.996	0.039	0.006
8	20,000	193.981	5	0.998	0.995	0.043	0.007
9	20,000	218.099	5	0.997	0.995	0.046	0.007
10	20,000	253.103	5	0.997	0.994	0.050	0.007
11	20,000	229.383	5	0.997	0.994	0.047	0.007
12	20,000	240.141	5	0.997	0.994	0.048	0.007

The EFA analyses were conducted using the factor analysis procedure in R, while the CFA analysis was conducted using R with the lavaan package Rosseel, 2012).

External Evidence: Relationship of Star Reading Scores to Scores on Other Tests of Reading Achievement

In an ongoing effort to gather evidence for the validity of Star Reading scores, continual research on score validity has been undertaken. In addition to original validity data gathered at the time of initial development, numerous other studies have investigated the correlations between Star Reading tests and other external measures. In addition to gathering concurrent validity estimates, predictive validity estimates have also been investigated. Concurrent validity was defined for students taking a Star Reading test and external measures within a two-month time period. Predictive validity provides an estimate of the extent to which scores on the Star Reading test predicted scores on criterion measures given at a later point in time, operationally defined as more than two months between the Star test (predictor) and the criterion test. Studies of Star Reading tests' concurrent and predictive correlations with other tests between 1999 and 2013 included the following other tests:

- AIMSweb
- Arkansas Augmented Benchmark Examination (AABE)
- California Achievement Test (CAT)
- Canadian Achievement Test (CAT)
- Colorado Student Assessment Program (CSAP)
- Comprehensive Test of Basic Skills (CTBS)
- Delaware Student Testing Program (DSTP)—Reading
- Dynamic Indicators of Basic Early Literacy Skills (DIBELS)—Oral Reading Fluency
- Florida Comprehensive Assessment Test (FCAT, FCAT 2.0)
- Gates-MacGinitie Reading Test (GMRT)
- Idaho Standards Achievement Test (ISAT)
- Illinois Standards Achievement Test—Reading
- Iowa Test of Basic Skills (ITBS)
- Kansas State Assessment Program (KSAP)

- Kentucky Core Content Test (KCCT)
- Metropolitan Achievement Test (MAT)
- Michigan Educational Assessment Program (MEAP)—English Language Arts and Reading
- Mississippi Curriculum Test (MCT2)
- Missouri Mastery Achievement Test (MMAT)
- New Jersey Assessment of Skills and Knowledge (NJ ASK)
- New York State Assessment Program
- North Carolina End-of-Grade (NCEOG) Test
- Ohio Achievement Assessment (OAA)
- Oklahoma Core Curriculum Test (OCCT)
- South Dakota State Test of Educational Progress (DSTEP)
- Stanford Achievement Test (SAT)
- State of Texas Assessments of Academic Readiness Standards Test (STAAR)
- Tennessee Comprehensive Assessment Program (TCAP)
- TerraNova
- Texas Assessment of Academic Skills (TAAS)
- Transitional Colorado Assessment Program (TCAP)
- West Virginia Educational Standards Test 2 (WESTEST 2)
- Woodcock Reading Mastery (WRM)
- Wisconsin Knowledge and Concepts Examination (WKCE)
- Wide Range Achievement Test 3 (WRAT 3)

Table 28 and Table 29 present summary evidence of concurrent validity collected between 1999 and 2013; between them, these tables summarize some 269 different analyses of concurrent validity with other tests, based on test scores of more than 300 thousand school children. The within-grade average concurrent validity coefficients for grades 1–6 varied from 0.72–0.80, with an overall average of 0.74. The within-grade average concurrent validity for grades 7–12 ranged from 0.65–0.76, with an overall average of 0.72.

Table 30 and Table 31 present summary evidence of predictive validity collected over the same time span: 1999 through 2013. These two tables display summaries of data 300 coefficients of correlation between Star Reading and other measures administered at points in time at least two months later than Star Reading; more than 1.45 million students' test scores are represented in these two tables. Predictive validity coefficients ranged from 0.69–0.72 in grades 1–6, with an average of 0.71. In grades 7–12 the predictive validity coefficients ranged from 0.72–0.87 with an average of 0.80.

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

For a compilation of all detailed validation information, see tables of correlations in "Appendix B: Detailed Evidence of Star Reading Validity".

Table 28: Concurrent Validity Data: Star Reading Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 1–6

Summary							
Grade(s)	**All**	**1**	**2**	**3**	**4**	**5**	**6**
Number of students	255,538	1,068	3,629	76,942	66,400	54,173	31,686
Number of coefficients	195	10	18		47	41	32
Average validity		0.80	0.73	0.72	0.72	0.74	0.72
Overall average	0.74						

Table 29: Concurrent Validity Data: Star Reading Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 7–12

Summary							
Grade(s)	**All**	**7**	**8**	**9**	**10**	**11**	**12**
Number of students	48,789	25,032	21,134	1,774	755	55	39
Number of coefficients	74	30	29	7	5	2	1
Average validity		0.74	0.73	0.65	0.76	0.70	0.73
Overall average	0.72						

Table 30: Predictive Validity Data: Star Reading Correlations (r) with External Tests Administered Fall 2005–Spring 2013, Grades 1–6

Summary							
Grade(s)	**All**	**1**	**2**	**3**	**4**	**5**	**6**
Number of students	1,227,887	74,887	188,434	313,102	289,571	217,416	144,477
Number of coefficients	194	6	10	49	43	47	39
Average validity		0.69	0.72	0.70	0.71	0.72	0.71
Overall average	0.71						

Table 31: Predictive Validity Data: Star Reading Correlations (r) with External Tests Administered Fall 2005–Spring 2013, Grades 7–12

Summary							
Grade(s)	**All**	**7**	**8**	**9**	**10**	**11**	**12**
Number of students	224,179	111,143	72,537	9,567	21,172	6,653	3,107
Number of coefficients	106	39	41	8	10	6	2
Average validity		0.72	0.73	0.81	0.81	0.87	0.86
Overall average	0.80						

Relationship of Star Reading Scores to Scores on State Tests of Accountability in Reading

The No Child Left Behind (NCLB) Act of 2001 required states to develop and employ their own accountability tests to assess students in ELA/Reading and Math in grades 3 through 8, and one high school grade. Until 2014, most states used their own accountability tests for this purpose. Renaissance Learning was able to obtain accountability test scores for many students who also took Star Reading; in such cases, it was feasible to calculate coefficients of correlation between Star Reading scores and the state test scores. Observed concurrent and predictive validity correlations are summarized below for the relationship between Star Reading and state accountability test scores for grades 3–8 for tests of both reading and language arts. Table 32 and Table 33 provide summaries from a variety of concurrent and predictive validity coefficients, respectively, for grades 3–8. Numerous state accountability tests have been used in this research.

Table 32: Concurrent Validity Data: Star Reading Correlations (r) with State Accountability Tests, Grades 3–8

Summary							
Grade(s)	**All**	**3**	**4**	**5**	**6**	**7**	**8**
Number of students	11,045	2,329	1,997	2,061	1,471	1,987	1,200
Number of coefficients	61	12	13	11	8	10	7
Average validity		0.72	0.73	0.73	0.71	0.74	0.73
Overall average	0.73						

Table 33: Predictive Validity Data: Star Reading Scaled Scores Predicting Later Performance for Grades 3–8 on Numerous State Accountability Tests

Summary							
Grade(s)	**All**	**3**	**4**	**5**	**6**	**7**	**8**
Number of students	22,018	4,493	2,974	4,086	3,624	3,655	3,186
Number of coefficients	119	24	19	23	17	17	19
Average validity		0.66	0.68	0.70	0.68	0.69	0.70
Overall average	0.68						

For Grades 3 to 8, Star Reading concurrent validity correlations by grade ranged between 0.71 to 0.74 with an overall average validity correlation of 0.71. For Grades 3 to 8, Star Reading predictive validity correlations by grade ranged between 0.66 to 0.70 with an overall average validity correlation of 0.68.

Relationship of Star Reading Scores to Scores on Multi-State Consortium Tests in Reading

In recent years, the National Governors' Association, in collaboration with the Council of Chief State School Officers (CCSSO), developed a proposed set of curriculum standards in English Language Arts and Math, called the Common Core State Standards. Forty-five states voluntarily adopted those standards; subsequently, many states have dropped them, but 14 states continue to use them or base their own state standards on them. Two major consortia were formed to develop assessments systems that embodied those standards: the Smarter Balanced Assessment Consortium (SBAC) and Partnership for Assessment of Readiness for College and Careers (PARCC). SBAC and PARCC end-of-year assessments have been administered in numerous states in place of those states' previous annual accountability assessments. Renaissance Learning was able to obtain SBAC and PARCC scores of many students who had taken Star Reading earlier in the same school years. Table 34 and Table 35, below, contain coefficients of correlation between Star Reading and the consortium tests.

Table 34: Concurrent Predictive Validity Data: Star Reading Scaled Scores Predicting Later Performance for Grades 3–8 on Smarter Balanced Assessment Consortium Test

Star Reading Predictive and Concurrent Correlations with Smarter Balanced Assessment Scores							
Grade(s)	**All**	**3**	**4**	**5**	**6**	**7**	**8**
Number of students	3,539	709	690	697	567	459	417
Fall Predictive		0.78	0.78	0.76	0.77	0.79	0.80
Winter Predictive		0.78	0.78	0.79	0.78	0.79	0.81
Spring Concurrent		0.79	0.82	0.80	0.70	0.79	0.81

Table 35: Concurrent and Predictive Validity Data: Star Reading Scaled Scores Correlations for Grades 3–8 with PARCC Assessment Consortium Test Scores

Star Reading Predictive and Concurrent Correlations with PARCC Assessment Scores							
Grade(s)	**All**	**3**	**4**	**5**	**6**	**7**	**8**
Number of students	22,134	1770	3950	3843	4370	4236	3965
Predictive Concurrent		0.82	0.85	0.82	0.81	0.83	0.80
Concurrent		0.83	0.82	0.78	0.79	0.80	0.77

The average of the concurrent correlations was approximately 0.79 for SBAC and 0.80 for PARCC. The average predictive correlation was 0.78 for the SBAC assessments, and 0.82 for PARCC.

Meta-Analysis of the Star Reading Validity Data

Meta-analysis is a statistical procedure for combining results from different sources or studies. When applied to a set of correlation coefficients that estimate test validity, meta-analysis combines the observed correlations and sample sizes to yield estimates of overall validity. In addition, standard errors and confidence intervals can be computed for overall validity estimates as well as within-grade validity estimates. To conduct a meta-analysis of the Star Reading validity data, 789 correlations reported in the *Star Reading Technical Manual* were combined and analyzed using a fixed-effects model for meta-analysis (see Hedges and Olkin, 1985, for a methodology description).

The results are displayed in Table 36. The table lists correlations within each grade, as well as results from combining data from all twelve grades. For each set of results, the table gives an estimate of the true validity, a standard error, and the lower and upper limits of a 95 percent confidence interval for the expected validity coefficient. Using the 789 correlation coefficients, the overall estimate of the validity of Star Reading is 0.79, with a standard error of 0.001. The 95 percent confidence interval allows one to conclude that the true validity coefficient for Star Reading is approximately 0.79. The probability of observing the 789 correlations reported in Table 28 through Table 35 if the true validity were zero, would be virtually zero. Because the 789 correlations were obtained with widely different tests, and among students from twelve different grades, these results provide strong support for the validity of Star Reading as a measure of reading skills.

Table 36: Results of the Meta-Analysis of Star Reading Correlations with Other Tests

Grade	Effect Size		95% Confidence Level		Total Correlations	Total N
	Validity Estimate	Standard Error	Lower Limit	Upper Limit		
1	0.70	0.00	0.69	0.70	18	78,022
2	0.78	0.00	0.78	0.78	32	196,114
3	0.79	0.00	0.79	0.79	131	628,336
4	0.79	0.00	0.79	0.79	125	594,712
5	0.79	0.00	0.79	0.79	123	518,411
6	0.79	0.00	0.79	0.79	106	330,475
7	0.79	0.00	0.79	0.79	98	276,218
8	0.79	0.00	0.79	0.79	98	225,704
9	0.78	0.01	0.78	0.79	19	27,952
10	0.82	0.01	0.81	0.82	21	34,913
11	0.74	0.01	0.73	0.74	15	32,798
12	0.86	0.02	0.85	0.87	3	3,146
All Grades	0.79	0.00	0.79	0.79	789	2,946,801

Additional Validation Evidence for Star Reading

This section provides summaries of new validation data along with tables of results. Data from four sources are presented here. They include a predictive validity study, a longitudinal study, a concurrent validity study in England, and a study of Star Reading's construct validity as a measure of reading comprehension.

A Longitudinal Study: Correlations with SAT9

Sadusky and Brem (2002) conducted a study to determine the effects of implementing Reading Renaissance (RR)[1] at a Title I school in the southwest from 1997–2001. This was a retrospective longitudinal study. Incidental to the study, they obtained students' Star Reading posttest scores and SAT9 end-of-year

1. Reading Renaissance is a supplemental reading program that uses Star Reading and Accelerated Reader.

Total Reading scores from each year and calculated correlations between them. Students' test scores were available for multiple years, spanning grades 2–6. Data on gender, ethnic group, and Title I eligibility were also collected.

Table 37 displays the observed correlations for the overall group. Table 38 displays the same correlations, broken out by ethnic group.

Overall correlations by year ranged from 0.66–0.73. Sadusky and Brem concluded that "Star results can serve as a moderately good predictor of SAT9 performance in reading."

Enough Hispanic and white students were identified in the sample to calculate correlations separately for those two groups. Within each ethnic group, the correlations were similar in magnitude, as Table 38 shows. This supports the assertion that Star Reading is valid for multiple student ethnicities.

Table 37: Correlations of the Star Posttest with the SAT9 Total Reading Scores 1998–2002[a]

Year	Grades	N	Correlation
1998	3–6	44	0.66
1999	2–6	234	0.69
2000	2–6	389	0.67
2001	2–6	361	0.73

a. All correlations significant, $p < 0.001$.

Table 38: Correlations of the Star Posttest with the SAT9 Total Reading Scores, by Ethnic Group, 1998–2002[a]

		Hispanic		White	
Year	Grade	N	Correlation	N	Correlation
1998	3–6	7 (n.s.)	0.55	35	0.69
1999	2–6	42	0.64	179	0.75
2000	2–6	67	0.74	287	0.71
2001	2–6	76	0.71	255	0.73

a. All correlations significant, $p < 0.001$, unless otherwise noted.

Concurrent Validity: An International Study of Correlations with Reading Tests in England

NFER, the National Foundation for Educational Research, conducted a study of the concurrent validity of both Star Reading and Star Math in 16 schools in England in 2006 (Sewell, Sainsbury, Pyle, Keogh and Styles, 2007). English primary and secondary students in school years 2–9 (equivalent to US grades

1–8) took both Star Reading and one of three age-appropriate forms of the Suffolk Reading Scale 2 (SRS2) in the fall of 2006. Scores on the SRS2 included traditional scores, as well as estimates of the students' Reading Age (RA), a scale that is roughly equivalent to the Grade Equivalent (GE) scores used in the US. Additionally, teachers conducted individual assessments of each student's attainment in terms of curriculum levels, a measure of developmental progress that spans the primary and secondary years in England.

Correlations with all three measures are displayed in Table 39, by grade and overall. As the table indicates, the overall correlation between Star Reading and Suffolk Reading Scaled Scores was 0.91, the correlation with Reading Age was 0.91, and the correlation with teacher assessments was 0.85. Within-form correlations with the SRS ability estimate ranged from 0.78–0.88, with a median correlation of 0.84, and ranged from 0.78–0.90 on Reading Age, with a median of 0.85.

Table 39: Correlations of Star Reading with Scores on the Suffolk Reading Scale and Teacher Assessments in a Study of 16 Schools in England

School Years[a]	Suffolk Reading Scale				Teacher Assessments	
	Test Form	N	SRS Score[b]	Reading Age	N	Assessment Levels
2–3	SRS1A	713	0.84	0.85	n/a	n/a
4–6	SRS2A	1,255	0.88	0.90	n/a	n/a
7–9	SRS3A	926	0.78	0.78	n/a	n/a
Overall		2,694	0.91	0.91	2,324	0.85

a. UK school year values are 1 greater than the corresponding US school grade. Thus, Year 2 corresponds to Grade 1, etc.
b. Correlations with the individual SRS forms were calculated with within-form raw scores. The overall correlation was calculated with a vertical Scaled Score.

Construct Validity: Correlations with a Measure of Reading Comprehension

The Degrees of Reading Power (DRP) test is widely recognized as a measure of reading comprehension. Yoes (1999) conducted an analysis to link the Star Reading Rasch item difficulty scale to the item difficulty scale of DRP. As part of the study, nationwide samples of students in grades 3, 5, 7, and 10 took two tests each (leveled forms of both the DRP and of Star Reading calibration tests). The forms administered were appropriate to each student's grade level. Both tests were administered in paper-and-pencil format. All Star Reading test forms consisted of 44 items, a mixture of vocabulary-in-context and extended passage comprehension item types. The grade 3 DRP test form (H-9) contained 42 items and all remaining grades (5, 7, and 10) consisted of 70 items on the DRP test.

Star Reading and DRP test score data were obtained on 273 students at grade 3, 424 students at grade 5, 353 students at grade 7, and 314 students at grade 10.

Item-level factor analysis of the combined Star and DRP response data indicated that the tests were essentially measuring the same construct at each of the four grades. Eigenvalues from the factor analysis of the tetrachoric correlation matrices tended to verify the presence of an essentially unidimensional construct. In general, the eigenvalue associated with the first factor was very large in relation to the eigenvalue associated with the second factor. Overall, these results confirmed the essential unidimensionality of the combined Star Reading and DRP data. Since DRP is an acknowledged measure of reading comprehension, the factor analysis data support the claim that Star Reading likewise measures reading comprehension.

Subsequent to the factor analysis, the Star Reading item difficulty parameters were transformed to the DRP difficulty scale, so that scores on both tests could be expressed on a common scale. Star Reading scores on that scale were then calculated using the methods of Item Response Theory. Table 40 below shows the correlations between Star Reading and DRP reading comprehension scores overall and by grade.

Table 40: Correlations between Star Reading and DRP Test Scores, Overall and by Grade

Grade	Sample Size	Test Form		Number of Items		Correlation
		Star Calibration	DRP	Star	DRP	
3	273	321	H-9	44	42	0.84
5	424	511	H-7	44	70	0.80
7	353	623	H-6	44	70	0.76
10	314	701	H-2	44	70	0.86
Overall	1,364					0.89

In summary, using item factor analysis Yoes (1999) showed that Star Reading items measure the same underlying construct as the DRP: reading comprehension. The overall correlation of 0.89 between the DRP and Star Reading test scores corroborates that. Furthermore, correcting that correlation coefficient for the effects of less than perfect reliability yields a corrected correlation of 0.96. Thus, both at the item level and at the test score level, Star Reading was shown to measure essentially the same construct as the DRP.

Investigating Oral Reading Fluency and Developing the Estimated Oral Reading Fluency Scale

During the fall of 2007 and winter of 2008, 32 schools across the United States that were then using both Star Reading and DIBELS oral reading fluency (DORF) for interim assessments participated in a research study to evaluate the relationship of Star Reading scores to oral reading fluency. Below are highlights of the methodology and results of the study.

A single-group design provided data for both evaluation of concurrent validity and the linking of the two score scales. For the linking analysis, an equipercentile methodology was used. Analysis was done independently for each of grades 1–4. To evaluate the extent to which the linking accurately approximated student performance, 90 percent of the sample was used to calibrate the linking model, and the remaining 10 percent were used for cross-validating the results. The 10 percent were chosen by a simple random function.

The 32 schools in the sample came from 9 states: Alabama, Arizona, California, Colorado, Delaware, Illinois, Michigan, Tennessee, and Texas. This represented a broad range of geographic areas, and resulted in a large number of students (N = 12,220). The distribution of students by grade was as follows:

- 1st grade: 2,001
- 2nd grade: 4,522
- 3rd grade: 3,859
- 4th grade: 1,838

The sample was composed of 61 percent of students of European ancestry; 21 percent of African ancestry; 11 percent of Hispanic ancestry; with the remaining 7 percent of Native American, Asian, or other ancestry.

Students were individually assessed using the DORF (DIBELS Oral Reading Fluency) benchmark passages. The students read the three benchmark passages under standardized conditions. The raw score for passages was computed as the number of words read correctly within the one-minute limit (WCPM, Words Correctly read Per Minute) for each passage. The final score for each student was the median WCPM across the benchmark passages, and was the score used for analysis. Each student also took a Star Reading assessment within two weeks of the DORF assessment.

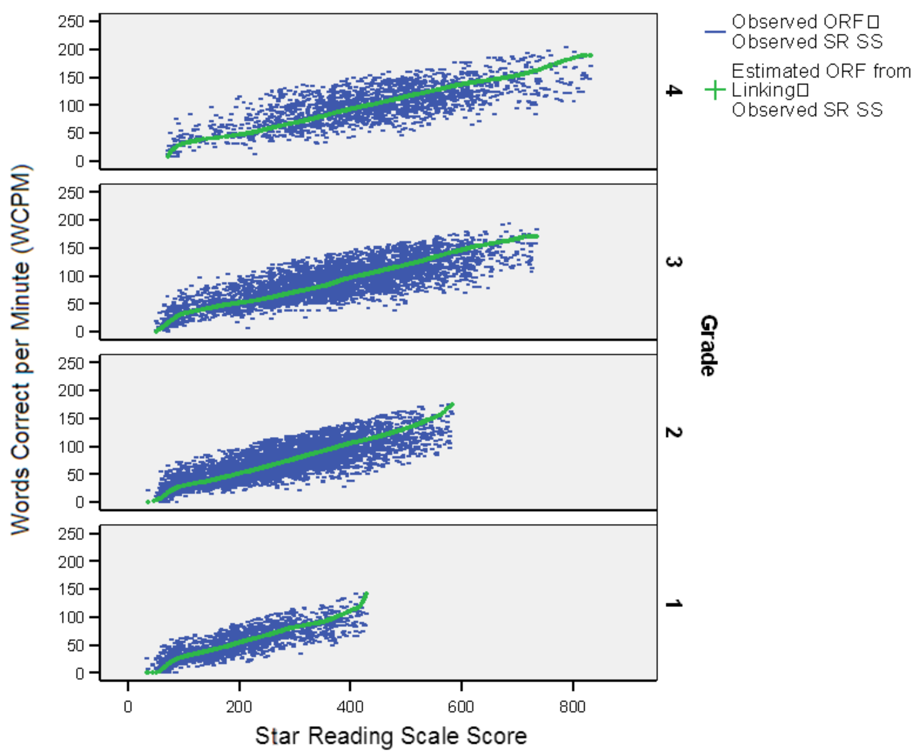
Descriptive statistics for each grade in the study on Star Reading Scaled Scores and DORF WCPM (words correctly read per minute) are found in Table 41.

Correlations between the Star Reading Scaled Score and DORF WCPM at all grades were significant ($p < 0.01$) and diminished consistently as grades increased. Figure 5 visualizes the scatterplot of observed DORF WCPM and SR Scaled Scores, with the equipercentile linking function overlaid. The equipercentile linking function appeared linear; however, deviations at the tails of the distribution for higher and lower performing students were observed. The root mean square errors of linking for grades 1–4 was found to be 14, 19, 22, and 25 WCPM, respectively.

Table 41: Descriptive Statistics and Correlations between Star Reading Scale Scores and DIBELS Oral Reading Fluency for the Calibration Sample

Grade	N	Star Reading Scale Score		DORF WCPM		Correlation
		Mean	SD	Mean	SD	
1	1,794	172.90	98.13	46.05	28.11	0.87
2	4,081	274.49	126.14	72.16	33.71	0.84
3	3,495	372.07	142.95	90.06	33.70	0.78
4	1,645	440.49	150.47	101.43	33.46	0.71

Figure 5: Scatterplot of Observed DORF WCPM and SR Scale Scores for Each Grade with the Grade Specific Linking Function Overlaid



Cross-Validation Study Results

The 10 percent of students randomly selected from the original sample were used to provide evidence of the extent to which the models based on the calibration samples were accurate. The cross-validation sample was intentionally kept out of the calibration of the linking estimation, and the results of the calibration sample linking function were then applied to the cross-validation sample.

Table 42 provides descriptive information on the cross-validation sample. Means and standard deviations for DORF WCPM and Star Reading Scaled Score for each grade were of a similar magnitude to the calibration sample. Table 43 provides results of the correlation between the observed DORF WCPM scores and the estimated WCPM from the equipercentile linking. All correlations were similar to results in the calibration sample. The average differences between the observed and estimated scores and their standard deviations are reported in Table 43 along with the results of one sample t-test evaluating the plausibility of the mean difference being significantly different from zero. At all grades the mean differences were not significantly different from zero, and standard deviations of the differences were very similar to the root mean square error of linking from the calibration study.

Table 42: Descriptive Statistics and Correlations between Star Reading Scale Scores and DIBELS Oral Reading Fluency for the Cross-Validation Sample

Grade	N	Star Reading Scale Score		DORF WCPM	
		Mean	SD	Mean	SD
1	205	179.31	100.79	45.61	26.75
2	438	270.04	121.67	71.18	33.02
3	362	357.95	141.28	86.26	33.44
4	190	454.04	143.26	102.37	32.74

Table 43: Correlation between Observed WCPM and Estimated WCPM Along with the Mean and Standard Deviation of the Differences between Them

Grade	N	Correlation	Mean Difference	SD Difference	t-test on Mean Difference
1	205	0.86	–1.62	15.14	t(204) = –1.54, p = 0.13
2	438	0.83	0.23	18.96	t(437) = 0.25, p = 0.80
3	362	0.78	–0.49	22.15	t(361) = –0.43, p = 0.67
4	190	0.74	–1.92	23.06	t(189) = –1.15, p = 0.25

Classification Accuracy of Star Reading

Accuracy for Predicting Proficiency on a State Reading Assessment

Star Reading test scores have been linked statistically to numerous state reading assessment scores. The linked values have been employed to use Star Reading to predict student proficiency in reading on those state tests. One example of this is a linking study conducted using a multi-state sample of students' scores on the PARCC consortium assessment[2]. The table below presents classification accuracy statistics for grades 3 through 8.

Table 44: Classification diagnostics for predicting students' reading proficiency on the PARCC consortium assessment from earlier Star Reading scores

Measure	Grade					
	3	4	5	6	7	8
Overall classification accuracy	86%	87%	86%	86%	86%	83%
Sensitivity	64%	73%	73%	69%	73%	70%
Specificity	93%	93%	90%	91%	91%	89%
Observed proficiency rate (OPR)	26%	29%	27%	24%	28%	29%
Projected proficiency rate (PPR)	22%	26%	26%	23%	27%	28%
Proficiency status projection error	–5%	–3%	0%	–1%	–1%	–1%
Area under the ROC curve	0.91	0.93	0.91	0.92	0.92	0.90

As the table shows, classification accuracy ranged from 83 to 87%, depending on grade. Area Under the Curve (AUC) was at least 0.90 for all grades. Specificity was especially high, and the projected proficiency rates were very close to the observed proficiency rates at all grades.

Numerous other reports of linkages between Star Reading and state accountability tests have been conducted. Reports are available at http://research.renaissance.com/.

Accuracy for Identifying At-Risk Students

In many settings, Star Reading is used to identify students considered "at risk" for reading difficulties requiring intervention, for example long in advance of state accountability assessment that will be used to classify students at the end of the school year. This section summarizes two studies done to evaluate the validity of cut scores based on Star Reading as predictors of "at risk" status later in the school year. In such cases, correlation coefficients are of less interest than classification

2. Renaissance Learning (2016). Relating Star Reading™ and Star Math™ to the Colorado Measure of Academic Success (CMAS) (PARCC Assessments) performance.

accuracy statistics, such as overall accuracy of classification, sensitivity and specificity, false positives and false negatives, positive and negative predictive power, receiver operating characteristic (ROC) curves, and a summary statistic called AUC (Area Under the Curve).[3] Summaries of the methodology and results of the two studies are given below.

Brief Description of the Current Sample and Procedure

Initial Star Reading classification analyses were performed using state assessment data from Arkansas, Delaware, Illinois, Michigan, Mississippi, and Kansas. Collectively these states cover most regions of the country (Central, Southwest, Northeast, Midwest, and Southeast). Both the Classification Accuracy and Cross Validation study samples were drawn from an initial pool of 79,045 matched student records covering grades 2–11.

A secondary analysis using data from a single state assessment was then performed. The sample used for this analysis was 42,771 matched Star Reading and South Dakota Test of Education Progress records of students in grades 3–8.

An ROC analysis was used to compare the performance data on Star Reading to performance data on state achievement tests, with “at risk” identification as the criterion. The Star Reading Scaled Scores used for analysis originated from assessments 3–11 months before the state achievement tests were administered. Selection of cut scores was based on the graph of sensitivity and specificity versus the Scaled Score. For each grade, the Scaled Score chosen as the cut point was equal to the score where sensitivity and specificity intersected. The classification analyses, cut points and outcome measures are outlined in Table 45. Area Under the Curve (AUC) values were all greater than 0.80. Descriptive notes for other values represented in the table are provided in the table footnote.

3. For descriptions of ROC curves, AUC, and related classification accuracy statistics, refer to Pepe, Janes, Longton, Leisenring, & Newcomb (2004) and Zhou, Obuchowski & Obushcowski (2002).

Table 45: Classification Accuracy in Predicting Proficiency on State Achievement Tests in Seven States[a]

Statistic[b]	Initial Analysis		Secondary Analysis	
	Value		Value	
False Positive Rate	21%		18%	
False Negative Rate	23%		22%	
Sensitivity	76%		78%	
Specificity	76%		82%	
Overall Classification Rate	76%		81%	
	Grade	**AUC**	**Grade**	**AUC**
AUC (ROC)	2	0.816		
	3	0.839	3	0.869
	4	0.850	4	0.882
	5	0.841	5	0.881
	6	0.833	6	0.883
	7	0.829	7	0.896
	8	0.843	8	0.879
	9	0.847		
	10	0.858		
	11	0.840		
	10	777		
	11	1,055		

a. Arkansas, Delaware, Illinois, Kansas, Michigan, Mississippi, and South Dakota.
b. The false positive rate is equal to the proportion of students incorrectly labeled “at-risk.” The false negative rate is equal to the proportion of students incorrectly labeled not “at-risk.” Likewise, sensitivity refers to the proportion of correct positive predictions while specificity refers to the proportion of negatives that are correctly identified (e.g., student will not meet a particular cut score).

Disaggregated Validity and Classification Data

In some cases, there is a need to verify that tests, such as Star Reading, as valid for different demographic groups. For that purpose, the data must be disaggregated, and separate analyses performed for each group. Table 46 shows the disaggregated classification accuracy data for ethnic subgroups and also the disaggregated validity data.

Table 46: Disaggregated Classification and Validity Data

Classification Accuracy in Predicting Proficiency on State Achievement Tests in 6 States (Arkansas, Delaware, Illinois, Kansas, Michigan, and Mississippi): by Race/Ethnicity										
	White, non-Hispanic (n = 17,567)		Black, non-Hispanic (n = 8,962)		Hispanic (n = 1,382)		Asian/Pacific Islander (n = 231)		American Indian/ Alaska Native (n = 111)	
False Positive	31%		44%		36%		17%		12%	
False Negative Rate	38%		12%		12%		24%		41%	
Sensitivity	62%		88%		88%		76%		59%	
Specificity	87%		56%		64%		83%		88%	
Overall Classification Rate	81%		67%		73%		82%		78%	
AUC (ROC)	**Grade**	**AUC**	**Grade**	**AUC**	**Grade**	**AUC**	**Grade**	**AUC**	**Grade**	**AUC**
	2	n/a	2	0.500	2	n/a	2	n/a	2	n/a
	3	0.863	3	0.828	3	0.868	3	0.913	3	0.697
	4	0.862	4	0.823	4	0.837	4	0.869	4	0.888
	5	0.853	5	0.832	5	0.839	5	0.855	5	0.919
	6	0.849	6	0.806	6	0.825	6	0.859	6	0.846
	7	0.816	7	0.784	7	0.866	7	0.904	7	0.900
	8	0.850	8	0.827	8	0.812	8	0.961	8	1.000
	9	1.000	9	0.848	9	n/a	9	n/a	9	n/a
	10	0.875	10	0.831	10	0.833	10	n/a	10	n/a
	11	0.750	11	1.000	11	n/a	11	n/a	11	n/a

Evidence of Technical Accuracy for Informing Screening and Progress Monitoring Decisions

Many school districts use tiered models such as Response to Intervention (RTI) or Multi-Tiered Systems of Support (MTSS) to guide instructional decision making and improve outcomes for students. These models represent a more proactive, data-driven approach for better serving students as compared with prior decision-making practices, including processes to:

- Screen all students to understand where each is in the progression of learning in reading, math, or other disciplines
- Identify at-risk students for intervention at the earliest possible moment
- Intervene early for students who are struggling or otherwise at-risk of falling behind; and
- Monitor student progress in order to make decisions as to whether they are responding adequately to the instruction/intervention

Assessment data are central to both screening and progress monitoring, and Star Reading is widely used for both purposes. This chapter includes technical information about Star Reading's ability to accurately screen students according to risk and to help educators make progress monitoring decisions. Much of this information has been submitted to and reviewed by the Center on Response to Intervention https://rti4success.org/ and/or the National Center on Intensive Intervention https://intensiveintervention.org/, two technical assistance groups funded by the US Department of Education.

For several years running, Star Reading has enjoyed favorable technical reviews for its use in informing screening and progress monitoring decision by the CRTI and NCII, respectively. The most recent reviews by CRTI indicate that Star Reading has a "convincing" level of evidence (the highest rating awarded) in the core screening categories, including classification accuracy, reliability, and validity. CRTI also notes that the extent of the technical evidence is "Broad" (again, the highest rating awarded) and notes that not only is the overall evidence compelling, but there are disaggregated data as well that shows Star Reading works equally well among subgroups. The most recent reviews by NCII indicate that there is full "convincing" evidence of Star Reading's psychometric quality for progress monitoring purposes, including reliability, validity, reliability of the slope, and validity of the slope. Furthermore, they find fully "convincing" evidence that Star Reading is sufficiently sensitive to student growth, has adequate alternate forms, and provides data-based guidance to educators on end-of-year benchmarks and when an intervention should be changed, among other categories. Readers may find additional information on Star Reading on those sites and should note that the reviews are updated on a regular basis, as their review standards are adjusted and new technical evidence for Star Reading and other assessments are evaluated.

Screening

According to the Center on Response to Intervention, "Screening is conducted to identify or predict students who may be at risk for poor learning outcomes. Universal screening assessments are typically brief, conducted with all students at a grade level, and followed by additional testing or short-term progress monitoring to corroborate students' risk status."[4]

Most commonly, screening is conducted with all students at the beginning of the year and then another two to four times throughout the school year. Star Reading is widely used for this purpose. In this section, the technical evidence supporting its use to inform screening decisions is summarized.

4. https://rti4success.org/essential-components-rti/universal-screening

Organizations of RTI/MTSS experts such as the Center on Response to Intervention and the RTI Action Network[5] are generally consistent in how measurement tools should be evaluated for their appropriateness as screeners. Key categories include the following:

1. **Validity and reliability.** Data on Star Reading's reliability were presented in the "Reliability and Measurement Precision" chapter of this manual. A wide array of validity evidence has been presented in this chapter, above; detailed tables of correlational data can be found in "Appendix B: Detailed Evidence of Star Reading Validity".
2. **Practicality and efficiency.** Screening measures should not require much teacher or student time. Because most students can complete a Star Reading test in 15–20 minutes or less, and because it is group administered and scored automatically, Star Reading is an exceptionally efficient general outcomes measure for reading.
3. **Classification accuracy metrics** including sensitivity, specificity, and overall predictive accuracy. These are arguably the most important indicators, addressing the main purpose of screening: When a brief screening tool indicates a student either is or is not at risk of later reading difficulties, how often is it accurate, and what types of errors are made?

It is common to use high stakes indicators such as state summative assessments as criterion measures for classification accuracy evaluation. Star Reading is linked to virtually every state summative assessment in the US as well as the ACT and SAT college entrance exams. The statistical linking of the Star Reading scale with these other measures' scales, combined with Star Reading growth norms (discussed in the Norming chapter of this manual) empowers Star Reading reports and data extracts to make predictions throughout the school year about future student performance. These predictions inform educator screening decisions in schools using an RTI/MTSS framework. (Educators are also free to use norm-referenced scores such as Percentile Ranks to inform screening decisions.)

Star Reading's classification accuracy results from several recent predictive studies are summarized in Table 47. Each study evaluated the extent to which Star Reading accurately predicted whether a student achieved a specific performance level on another reading or English Language Arts measure. The specific performance level (cut point) varies by assessment and grade. Cut points are set by assessment developers and sponsors, which in the case of state summative exams usually means the state department of education and/or state board of education. State assessments generally have between three and five

5. http://www.rtinetwork.org/learn/research/universal-screening-within-a-rti-model

performance levels, and the cut point used in these analyses refers to the level the state has determined indicates meeting grade level reading or English Language Arts standards. For instance, the cut point on California's CAASPP is Level 3, also known as "Standard Met." On Louisiana's LEAP 2025 the cut point is at the "Mastery" level. In the case of ACT and SAT, the cut point established by the developers (ACT and College Board, respectively) indicates an estimated level of readiness for success in college.

Table 47: Summary of classification accuracy metrics from recent studies linking Star Reading with summative reading and English Language Arts measures

Assessment	Grade/s Covered	Date study completed	Study sample size	Average result across all grades			
				Overall classification accuracy	Sensitivity	Specificity	Area under ROC curve
ACT English (college readiness)	11	4/22/2016	14,248	80%	76%	82%	0.87
ACT Reading (college readiness)	11	4/22/2016	14,228	83%	62%	90%	0.86
ACT Aspire	3–10	6/1/2017	44,877	84%	81%	84%	0.92
California Assessment of Student Performance and Progress (CAASPP) (Smarter Balanced)	3–8	10/30/2015	51,835	84%	86%	82%	0.92
Florida Standards Assessments (FSA)	3–8	6/30/2015	41,178	84%	84%	83%	0.92
Georgia Milestones	3–8	7/1/2017	44,436	87%	79%	90%	0.94
Illinois Partnership for Assessment of Readiness for College and Careers (PARCC) Assessments	3–10	7/13/2016	27,415	86%	70%	91%	0.91
Louisiana Educational Assessment Program (LEAP 2025)	3–8	12/1/2017	33,815	84%	90%	69%	0.90
Maine Educational Assessment (MEA)	3–8	7/1/2017	945	83%	78%	86%	0.93
Mississippi Academic Assessment Program (MAAP)	3–8	2/1/2017	13,590	84%	80%	87%	0.92
Missouri Assessment Program (MAP) Grade-Level Tests	3–8	3/14/2017	30,626	85%	83%	87%	0.96

Table 47: Summary of classification accuracy metrics from recent studies linking Star Reading with summative reading and English Language Arts measures

Assessment	Grade/s Covered	Date study completed	Study sample size	Average result across all grades			
				Overall classification accuracy	Sensitivity	Specificity	Area under ROC curve
North Carolina READY End-of-Grade (EOG)	3–8	2/16/2015	396,075	81%	83%	78%	0.89
Ohio State Tests	3–8	12/20/2016	27,487	85%	83%	87%	0.93
Pennsylvania's System of School Assessment (PSSA)	3–8	12/19/2016	7,383	85%	91%	72%	0.92
SAT (college entrance)	11						
South Carolina College-and Career-Ready Assessments (SC READY)	3–8	12/5/2016	10,011	86%	85%	86%	0.94
State of Texas Assessments of Academic Readiness (STAAR)	3–8	7/1/2017	3,915	83%	71%	88%	0.90
Wisconsin Forward Exam	3–8	12/22/2016	39,605	88%	73%	93%	0.94

Notes:

- Some tests, such as the Smarter Balanced (indicated above for California) and PARCC (indicated above for Illinois) are used in multiple states, so those results may apply to other states not listed here.
- Overall classification accuracy refers to the percentage of correct classifications.
- Sensitivity refers to the rate at which Star Reading identifies students as being at-risk who demonstrate a poor learning outcome at a later point in time. Sensitivity can be thought of as the true positive rate. Screening tools with high sensitivity help ensure that students who truly need intervention will be identified to receive it.
- Specificity refers to the rate at which Star Reading identifies students as being not at-risk who perform satisfactorily at a later point in time. Specificity can be thought of as a true negative rate. Screening tools with high specificity help ensure that scarce resources are not invested in students who do not require extra assistance.
- Area under the ROC (Receiver Operating Characteristic) curve is a powerful indicator of overall accuracy. The ROC curve a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity) for the full range of possible screener (Star Reading) cut points. The area under ROC Curve (AUC) is an overall indication of the diagnostic accuracy of the curve. AUC values range between 0 and 1 with 0.5 indicating a chance level of accuracy. The Center for Response to Intervention considers results at or above 0.85 to be an indication of convincing evidence of classification accuracy.[6]

Note that many states tend to not use the same assessment system for more than a few consecutive years, and Renaissance endeavors to keep the Star Reading classification reporting as up to date as possible. Those interested in reviewing the full technical reports for these or other state assessments are encouraged to visit

6. https://rti4success.org/resources/tools-charts/screening-tools-chart/screening-tools-chart-rating-system

http://research.renaissance.com/advancedsearch.asp and search by state name for the Star Reading linking reports (e.g., “Wisconsin linking”).

Progress Monitoring

According to the National Center on Intensive Intervention, “progress monitoring is used to assess a student’s performance, to quantify his or her rate of improvement or responsiveness to intervention, to adjust the student’s instructional program to make it more effective and suited to the student’s needs, and to evaluate the effectiveness of the intervention.”[7]

In an RTI/MTSS context, progress monitoring involves frequent assessment—usually occurring once every 1–4 weeks—and often involves only those students who are receiving additional instruction after being identified as at-risk via the screening process. Ultimately, educators use progress monitoring data to determine whether a student is responding adequately to the instruction, or whether adjustments need to be made to the instructional intensity or methods. The idea is to get to a decision quickly, with as little testing as possible, so that valuable time is not wasted on ineffective approaches. Educators make these decisions by comparing their performance against a goal set by the educator. Goals should be “reasonable yet ambitious”[8] as recommended by Shapiro (2008), and Star Reading offers educators a variety of guidance to set normative or criterion-referenced goals that meet these criteria.

The RTI Action Network, National Center on Intensive Intervention, and other organizations offering technical assistance to schools implementing RTI/MTSS models are generally consistent in encouraging educators to select assessments for progress monitoring that have certain characteristics.

7. https://intensiveintervention.org/ncii-glossary-terms#ProgresMonitoring
8. Shapiro, E. S. (2008). Best practices in setting progress-monitoring goals for academic skill improvement. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 141-157). Bethesda, MD: National Association of School Psychologists.

A summary of those characteristics and relevant information about Star Reading is provided below.

1. **Evidence of psychometric quality.**

    a. **Reliability and validity.** Summaries of the available evidence supporting Star Reading's reliability and validity are presented in the chapter on "Reliability and Measurement Precision" and throughout this Validity chapter.

    b. **Reliability of the slope.** Because progress monitoring decisions often involve the student's rate of progress over multiple test administrations, the characteristics of the student's slope of improvement, or trend line, are also important. A study was conducted in 2017 by Renaissance Learning to evaluate reliability of slope for at-risk students who were being progress monitored during the 2016–17 school year. Specifically, the sample included 218,689 students who began the year below the 30th Percentile Rank in Star Reading and were assessed 10 or more times during the school year, with a minimum of 140 days between first and last test.

    Every student's Star Reading test records were sorted in chronological order. Each test record was coded as either an odd- or even-numbered test. Slopes were estimated for each student's odd-number tests and also for the even-numbered tests using ordinary least squares regression. Then, the odd and even slopes were correlated. The table below summarizes the Pearson correlation coefficients by grade, indicating a consistently strong association between even and odd numbered test slopes.

Table 48: Star Reading Reliability of the Slope Coefficients by grade, 1–12

Grade	n	Coefficient
1	14,179	0.76
2	43,978	0.93
3	52,670	0.94
4	37,862	0.93
5	31,326	0.93
6	16,990	0.94
7	9,683	0.94
8	7,786	0.94
9	2,483	0.94
10	1,549	0.94
11	799	0.94
12	384	0.95

2. **Produce a sufficient number of forms.** Because Star Reading is computer adaptive and its item bank comprises more than six thousand items, there are at a minimum, several hundred alternate forms for a student at a given ability level. This should be more than sufficient for even the most aggressive progress monitoring testing schedule.

   A variety of grade-specific evidence is available to demonstrate the extent to which Star Reading can reliably produce consistent scores across repeated administrations of the same or similar tests to the same individual or group. These include:

   a. Generic reliability, defined as the proportion of test score variance that is attributable to true variation in the trait or construct the test measures.

   b. Alternate forms reliability, defined as the correlation between test scores on repeated administrations to the same examinees.

   Grade-level results are summarized in the "Reliability and Measurement Precision" chapter.

3. **Practicality and efficiency.** As mentioned above, most students complete Star Reading in 15–20 minutes. It is auto-scored and can be group administered, requiring very little educator involvement, making it an efficient progress monitoring solution.

4. **Specify criterion for adequate growth and benchmarks for end-of-year performance levels.** Goal-setting decisions are handled by local educators, who know their students best and are familiar with the efficacy and intensity of the instructional supports that will be offered. That said, publishers of assessments used for progress monitoring are expected to provide empirically based guidance to educators on setting goals.

   Star Reading provides guidance to inform goal setting using a number of different metrics, including the following:

   a. **Student Growth Percentile.** SGP describes a student's velocity (slope) relative to a national sample of academic peers—those students in the same grade with a similar score history. SGPs work like Percentile Ranks (1–99 scale) but once an SGP goal has been set, it is converted to a Scaled Score goal at the end date specified by the teacher. An SGP-defined goal can be converted into an average weekly increase in a Scaled Score metric, if educators prefer to use that. Many teachers select either SGP 50 (indicating typical or expected growth) as minimum acceptable growth, or something indicating accelerated growth, such as 65 or 75. A helpful feature of SGP is that it can be used as a "reality check" for any goal, whether it be in an SGP metric or something else (e.g., Scaled Score, Percentile Rank). SGP estimates the likelihood that the student will achieve a level of growth or later performance. For example, a goal associated with an SGP of 75 indicates that only about 25 percent of the student's academic peers would be expected to achieve that level of growth.

b. **State test proficiency.** As described in the Screening section, the fact that Star Reading is linked to virtually every state assessment enables educators to select values on the Star scale that are approximately equivalent to states' defined proficiency level cut points for each grade.

c. **Percentile Rank and Scaled Score.** Educators may also enter custom goals using Percentile Rank or Scaled Score metrics.

Additional Research on Star Reading as a Progress Monitoring Tool

A study by Bulut & Cormier (2018) evaluated Star Reading as a progress monitoring tool, concluding:

- Although relatively little research exists on using computer adaptive measures for progress monitoring as opposed to curriculum based measurement probes, the study concluded it was possible to use Star Reading for progress monitoring purposes.
- Sufficiently reliable progress monitoring slopes could be generated in as few as five Star Reading administrations.
- The duration of Star Reading progress monitoring (i.e., over how many weeks) should be conducted is a function of the amount of typical growth by grade in relation to measurement error. For earlier grades (when student rates of growth are greatest), that amount of time could be as little as six weeks. For middle grades, 20 weeks should be sufficient.
- These two findings challenge popular rules of thumb about progress monitoring frequency and duration (most of which are derived from CBM probe studies), which often involve weekly testing over periods of time that are selected due to popular convention rather than empirical evidence.
- Using Theil-Sen regression procedures to estimate slope as opposed to OLS could reduce the influence of outlier scores, and thus provide a more accurate picture of student growth.

Differential Item Functioning

Ensuring that an assessment is not biased against different demographic subgroups that take the assessment is a fundamental aspect of showing test fairness and providing validity evidence to support the interpretations and uses of the assessment. One strategy that is often used as part of evaluating test fairness is a strategy known as differential item functioning (DIF). DIF occurs when two or more demographic subgroups perform differently on an item after controlling for performance on the test (Holland & Thayer, 1988; Zumbo, 2007). In other words, for students of similar ability, an item that displays DIF may appear to favor one group of students based on demographics such as gender and/or race/ethnicity.

There are many different methods that one can use to investigate items for DIF, including item response theory methods, observed score methods, and a variety of nonparametric approaches (Zumbo, 2007). The method that Star Reading uses to evaluate items for DIF is a method known as logistic regression (Rogers & Swaminathan, 1993; Swaminathan & Rogers, 1990; Swaminathan, 1994). With this approach, student item responses are regressed on student ability estimates from Star Reading as well as their subgroup membership and the student ability and subgroup membership interaction. To conduct a DIF analysis, a reference group and a focal group is defined. For instance, male is the reference group for gender while female is the focal group. Similarly, Caucasian is the reference group for race/ethnicity with the minority race/ethnic groups being focal groups. Separate models are run for DIF for male versus female, black versus white, Hispanic versus white, Asian versus white, and Native American versus white.

Items are flagged for DIF using a blended approach that employs a Chi-square test of statistical significance to determine if DIF is present and then assessing whether any evidence of DIF is practically significant using the Nagelkerke R^2 statistic (1991), a common effect size measure used in DIF investigations with logistic regression (Jodion & Gierl, 2001). Using the Nagelkerke R^2 statistic, items are categorized as exhibiting negligible DIF if the null hypothesis is not rejected or the R^2 statistic is less than 0.035, moderate DIF if the null hypothesis is rejected and the R^2 statistic is greater than or equal to 0.035 and less than 0.070, or large DIF if the null hypothesis is rejected and the R^2 statistic is greater than or equal to 0.070 (Jodion & Gierl, 2001).

There are a couple of points in the Star Reading assessment development cycle when items are evaluated for DIF. The first time point is when an item is included as a field test item as part of Star Reading's item calibration process. During item calibration, new assessment items are tried out with different groups of students to make sure that items have appropriate statistical and psychometric properties before they are used operationally and count towards a student's score. The second time point is when the full item bank of operational test items is recalibrated for scale maintenance to check whether the statistical and psychometric properties of the items have remained similar after the items become operational.

It is important to point out that just because an item is flagged for DIF against one or more subgroups does not necessarily mean that the item is biased. There are many possible explanations why an item may be statistically flagged for DIF. All items that are statistically flagged as having non-negligible DIF are marked for a bias and sensitivity review by the Content team. This review process consists of several subject matter experts with diverse perspectives and different backgrounds looking at and reviewing each item to see if there is any content in the item that may be biased against a particular subgroup and might explain why the item was statistically flagged for DIF. Items identified as being biased for any reason are removed from the item bank and do not appear on the Star Reading test. The statistical flagging of items for DIF as well as the bias and sensitivity review by the Content team helps ensure test fairness and that the items that appear on Star Reading do not favor any group of students that may take the test.

As shown in Figure 6 only 2% of about 6,000 items in the Star Reading item bank showed any evidence of DIF when Star Reading was recalibrated in 2021.

Figure 6: Summary of Star Reading Items with DIF

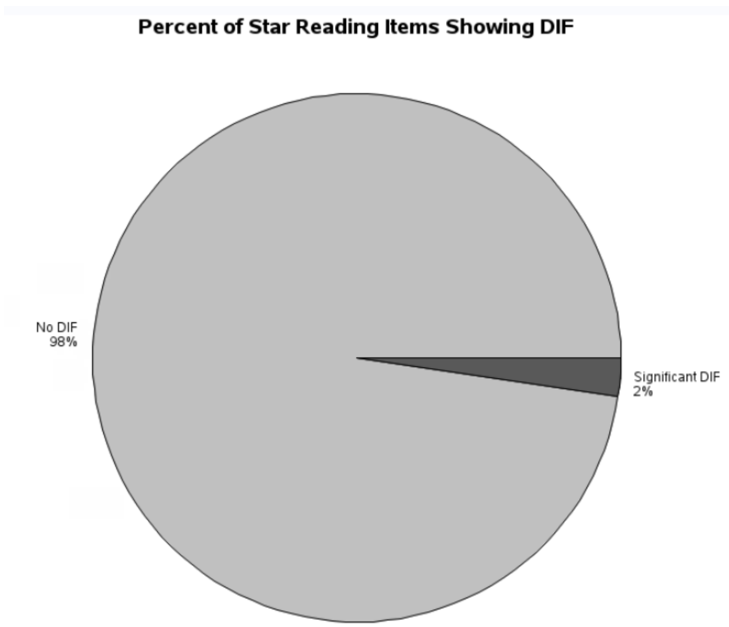


Table 49 shows the DIF results by reference and focal groups from various the DIF analyses. These results suggest that of the thousands of items analyzed very few items were flagged for DIF. There were 1.00% of items categorized with non-negligible DIF for male versus female, 0.04% of items flagged with non-negligible DIF for Asian versus white, 0.34% of items flagged with non-negligible DIF for black versus white, 0.94% of items flagged with non-negligible DIF for Hispanic versus white, and 0.00% of items flagged with non-negligible DIF for Native American versus white. These results provide evidence of the fairness of the Star Reading test for different demographic subgroups that take the assessment. As previously noted, any items that show DIF are removed from operational use.

Table 49: Percentage of Items for Different DIF Analyses

DIF Analysis	Percent of Items Showing DIF
Female versus Male	1.00%
Asian versus White	0.04%
Black versus White	0.34%
Hispanic versus White	0.94%
Native American versus White	0.00%

Summary of Star Reading Validity Evidence

The validity data presented in this technical manual includes evidence of Star Reading's concurrent, predictive, and construct validity, as well as classification accuracy statistics, and strong measures of association with non-traditional reading measures such as oral reading fluency. Exploratory and confirmatory factor analyses provided evidence that Star Reading measures a unidimensional construct, consistent with the assumption underlying its use of the Rasch model. The Meta-Analysis section showed the average uncorrected correlation between Star Reading and all other reading tests to be 0.79. (Many meta-analyses adjust the correlations for range restriction and attenuation to less than perfect reliability; had we done that here, the average correlation would have exceeded 0.85.) Correlations with specific measures of reading ability were often higher than this average. For example, Yoes (1999) found within-grade correlations with DRP averaging 0.81. When these data were combined across grades, the correlation was 0.89. The latter correlation may be interpreted as an estimate of the overall construct validity of Star Reading as a measure of reading comprehension. Yoes also reported that results of item factor analysis of DRP and Star Reading items yielded a single common unidimensional factor. This provides strong support for the claim that Star Reading is a measure of reading comprehension.

International data from the UK show even stronger correlations between Star Reading and widely used reading measures there: overall correlations of 0.91 with the Suffolk Reading Scale, median within-form correlations of 0.84, and a correlation of 0.85 with teacher assessments of student reading.

Finally, the data showing the relationship between the current, standards- based Star Reading Enterprise test and scores on specific state accountability tests and on the SBAC and PARCC Common Core consortium tests show that the correlations with these summative measures are consistent with the meta-analysis findings.

Norming

Two distinct kinds of norms are described in this chapter: test score norms and growth norms. The former refers to distributions of test scores themselves. The latter refers to distributions of changes in test scores over time; such changes are generally attributed to growth in the attribute that is measured by a test. Hence distributions of score changes over time may be called "growth norms."

Background

National norms for Star Reading version 1 were first collected in 1996. Substantial changes introduced in Star Reading version 2 necessitated the development of new norms in 1999. Those norms were used until new norms were developed in 2008. Since 2008, Star Reading norms have been updated three times (2014, 2017, and 2022). The 2022 norms went live in Star Reading in the 2022–2023 school year. This chapter describes the development of the 2022 norms.

From 1996 through mid-2011, Star Reading was primarily a measure of reading comprehension comprising short vocabulary-in-context items and longer passage comprehension items. The current version of Star Reading, introduced in June 2011, is a standards-based assessment that measures a wide variety of skills and instructional standards, as well as reading comprehension. To develop the current version of Star Reading, scale scores were equated to the scale used in earlier versions of Star Reading. The equating analyses demonstrated that, despite its distinctive content, the latent attribute underlying the current version is the same one underlying previous versions of Star Reading. It measures the same broad construct, and reports student performance on the same score scale. As part of the 2014 norming process, scores from the older version of Star Reading were equated to the current version. The 2022 norms are based on the current standards-based version of Star Reading.

The 2022 Star Reading Norms

Prior to development of the 2022 Star Reading norms, a new reporting scale was developed, called the Unified scale. The Unified scale is a linear transformation of Star Reading's Rasch ability scale to a new integer scale that is also used in Star Early Literacy. The Star Unified scale makes it possible to report performance on both of those Star assessments on the same scale.

The original Star Reading scale, now referred to as the "Enterprise" score scale, was based on a nonlinear transformation of Rasch scores. Both the Enterprise and the Unified scale scores will be available to Star test users during the planned transition to the Unified scale as the default reporting scale. The Unified scale is the default scale in the software for the 2022–2023 school year. This chapter displays normative summary data for both the Enterprise and Unified scales.

New U.S. norms for Star Reading assessments were introduced at the start of the 2017–18 school year. Separate early fall and late spring norms were developed for grades Kindergarten through 12. Before the introduction of the 2017 Star Reading norms, the reference populations for grades Kindergarten through 3 consisted only of students taking Star Reading; students who only took Star Early Literacy were excluded from the Star Reading norms, and vice versa. Consequently, previous Star Reading and Star Early Literacy norms for this grade range were not completely representative of the full range of literacy development in those grades. To address this, the concept of “Star Early Learning” was introduced. That concept acknowledges the overlap of literacy development content between the Star Reading and Early Literacy assessments, and encompasses in the normative reference group all students in each of grades K–3 who have taken either the Reading assessment, the Early Literacy assessment, or both.

The norms introduced in 2022 are based on test scores of K–3 students taking either the Reading assessment, or the Early Literacy one, or both. These norms are based on the use of the Unified scale, which allowed performance on both Star Early Literacy and Star Reading to be measured on the same scale. A new feature of the 2022 norms are norms for students in pre-K for Star Early Literacy, which are based on students taking Star Early Literacy in this grade. Pre-K norms are not available for Star Reading because students do not typically take Star Reading in this grade.

Due to testing impacts from COVID-19, the 2022 Star Reading norms are based on Star Reading and Star Early Literacy test data collected over the course of the 2018–2019 school year. Students participating in the norming study took assessments between August 1, 2018, and June 30, 2019. Students took the Star Reading or Early Literacy tests under normal test administration conditions. No specific norming test was developed and no deviations were made from the usual test administration. Thus, students in the norming sample took Star Reading or Star Early Literacy tests as they are administered in everyday use.

Sample Characteristics

During the norming period, a total of 1,223,730 US students in grades Pre-K–3 took Star Early Literacy while 9,665,081 US students in grades K–12 took Star Reading tests, using Renaissance servers hosted by Renaissance Learning. The first step in sampling was to select representative fall and spring student samples:

Students who had tested in the fall, in the spring, or in both the fall and spring of the 2018–2019 school year. From the fall and the spring samples, stratified subsamples were randomly drawn based on student grade and ability decile. The grade and decile sampling was necessary to ensure adequate and similar numbers of students in each grade, and each decile within grade.

Because these norming data were convenience samples drawn from the Star Reading and Star Early Literacy customer base, steps were taken to ensure the resulting norms were nationally representative of grades K–12 US student populations with regard to certain important characteristics. A post-stratification procedure was used to adjust the sample's proportions to the approximate national proportions on three key variables: geographic region, district socio-economic status, and district/school size. These three variables were chosen because they had previously been used in Star norming studies to draw nationally representative samples, are known to be related to test scores, and were readily available for the schools in the Renaissance hosted database.

The final norming sample size for grades K–12, after selecting only students with test scores in either the fall or the spring or both fall and spring in the norming year, and further sampling by grade and ability decile was 4,703,786 students. There were 3,766,400 students in the fall norming sample and 2,151,320 students in the spring norming sample; 1,213,934 students were included in both norming samples. These students came from 24,295 schools across the 50 states and the District of Columbia.

Table 50 and Table 51 provide a breakdown of the number of students participating per grade in the fall and in the spring, respectively.

Table 50: Numbers of Students per Grade in the Fall Norms Sample

Grade	N	Grade	N	Grade	N	Grade	N
K	219,980	4	624,810	8	175,210	12	23,420
1	373,300	5	502,860	9	57,120	Total	3,766,400
2	563,280	6	340,390	10	67,800		
3	566,780	7	213,030	11	38,420		

Table 51: Numbers of Students per Grade in the Spring Norms Sample

Grade	N	Grade	N	Grade	N	Grade	N
K	204,950	4	341,290	8	61,530	12	4,510
1	247,430	5	264,970	9	41,130	Total	2,151,320
2	325,560	6	161,650	10	33,110		
3	351,450	7	102,030	11	11,710		

National estimates of US student population characteristics were obtained from two entities: the National Center for Educational Statistics (NCES) and Market Data Retrieval (MDR).

- National population estimates of students' demographics (ethnicity and gender) in grades K–12 were obtained from NCES; these estimates were from the 2017–2018 school year for private schools and the 2018–2019 school year for public schools, the most recent data available. National estimates of race/ethnicity were computed using the NCES data based on single race/ethnicity and also a multiple race category. The NCES data reflect the most recent census data from the US census bureau.
- National estimates of school-related characteristics were obtained from October 2018 Market Data Retrieval (MDR) information. The MDR database contains the most recent data on schools, some of which may not be reflected in the NCES data.

Table 52 on page 101 shows national percentages of children in grades K–12 by region, school/district enrollment, district socio-economic status, location, and school type (public versus non-public), along with the corresponding percentages in the fall and in the spring norming samples. MDR estimates of geographic region were based on the four broad areas identified by the National Educational Association as Northeastern, Midwestern, Southeastern, and Western regions. The specific states in each region are shown below.

Geographic Region

Using the categories established by the National Center for Education Statistics (NCES), students were grouped into four geographic regions as defined below: Northeast, Southeast, Midwest, and West.

Northeast:

Connecticut, District of Columbia, Delaware, Massachusetts, Maryland, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont

Southeast:

Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia

Midwest:

Iowa, Illinois, Indiana, Kansas, Minnesota, Missouri, North Dakota, Nebraska, Ohio, South Dakota, Michigan, Wisconsin

West:

Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, New Mexico, Nevada, Oklahoma, Oregon, Texas, Utah, Washington, Wyoming

School Size

Based on total school enrollment, schools were classified into one of three school size groups: small schools had under 200 students enrolled, medium schools had between 200–499 students enrolled, and large schools had 500 or more students enrolled.

Socioeconomic Status as Indexed by the Percent of School Students with Free and Reduced Lunch

Schools were classified into one of four classifications based on the percentage of students in the school who had free or reduced student lunch. The classifications were coded as follows:

- High socioeconomic status (0%–24%)
- Above-median socioeconomic status (25%–49%)
- Below-median socioeconomic status (50%–74%)
- Low socioeconomic status (75%–100%)

No students were sampled from the schools that did not report the percent of school students with free and reduced lunch.

The norming sample also included private schools, Catholic schools, students with disabilities, and English Language Learners as described below.

Table 52: Sample Characteristics Along with National Population Estimates and Sample Estimates

		National Estimates	Fall Norming Sample	Spring Norming Sample
Region	Midwest	21.0%	18.8%	19.1%
	Northeast	18.6%	10.1%	15.8%
	Southeast	25.0%	27.6%	19.9%
	West	35.4%	43.6%	45.2%
School Enrollment	< 200	3.7%	4.1%	4.4%
	200–499	27.9%	37.7%	39.6%
	≥ 500	68.4%	58.2%	56.0%
District Socioeconomic Status	Low	20.7%	24.2%	25.1%
	Below Median	21.5%	23.1%	21.3%
	Above Median	24.4%	23.2%	22.3%
	High	33.5%	29.5%	31.4%
Location	Rural	14.4%	20.0%	17.3%
	Suburban	41.7%	36.8%	39.3%
	Town	11.4%	15.2%	14.0%
	Urban	32.5%	28.0%	29.5%
School Type	Public	91.9%	91.6%	90.6%
	Non-Public	8.1%	8.4%	9.4%

Table 53 provides information on the demographic characteristics of students in the sample and national percentages provided by NCES. No weighting was done on the basis of these demographic variables; they are provided to help describe the sample of students and the schools they attended. Because Star assessment users do not universally enter individual student demographic information such as gender and ethnicity/race, some students were missing demographic data; the sample summaries in Table 53 are based on only those students for whom gender and ethnicity information were available. In addition to the student demographics shown, an estimated 7.4% of the students in the norming sample were gifted and talented (G&T)[1] as approximated by the 2011–2012 school data collected by the Office of Civil Rights (OCR). OCR is a subsidiary of the US Department of Education. School type was defined to be either public (including charter schools) or non-public (private, Catholic).

1. This estimate is based on data from the previous version of Star Reading norms. Given the similarity of the user pools for those and the 2017 norms, the current percentage is expected to be approximately the same.

Table 53: Student Demographics and School Information: National Estimates and Sample Percentages

			National Estimate	Fall Norming Sample	Spring Norming Sample
Gender	Public	Female	48.7%	50.1%	49.5%
		Male	51.3%	49.9%	50.5%
	Non-Public	Female	–	51.1%	50.6%
		Male	–	48.9%	49.4%
Race/Ethnicity	Public	American Indian	1.0%	1.5%	1.6%
		Asian	5.6%	5.9%	5.4%
		Black	15.1%	15.0%	17.9%
		Hispanic	27.1%	25.2%	22.7%
		White	47.1%	49.0%	52.4%
		Multiple Race[a]	4.0%	3.2%	2.6%
	Non-Public	American Indian	0.6%	0.6%	0.4%
		Asian	7.2%	11.1%	11.0%
		Black	9.2%	4.7%	5.4%
		Hispanic	11.5%	30.0%	32.0%
		White	66.7%	46.3%	43.5%
		Multiple Race[a]	4.9%	7.4%	7.6%

a. Students identified as belonging to two or more races.

Test Administration

All students took the current version Star Reading or Early Literacy tests under normal administration procedures. Some students in the normative sample took the assessment two or more times within the norming windows; scores from their initial test administration in the fall and the last test administration in the spring were used for computing the norms.

Data Analysis

Student test records were compiled from the complete database of Star Reading and Early Literacy Renaissance Place users. Data were from a single school year from August 2018 to June 2019. Students' Unified scale Rasch scores on their first Star Reading or Early Literacy test taken during the first and second months of the school year based on grade placement were used to compute norms for the fall; students' Rasch scores on the last Star Reading or Early Literacy test taken during the 8th and 9th months of the school year were used to compute norms for the spring. Interpolation was used to estimate norms for times of the year between the first month in the fall and the last month in the spring. The norms were based on the distribution of Rasch scores for each grade.

As noted above, a post-stratification procedure was used to approximate the national proportions on key characteristics. Post stratification weights from the regional, district socio-economic status, and school size strata were computed and applied to each student's Unified Rasch ability estimate. Norms were developed based on the weighted Rasch ability estimates and then transformed to Unified as well as Enterprise Star Reading scaled scores.[2] Table 54 provides descriptive statistics for each grade with respect to the normative sample performance, in the Unified scaled score units. Table 55 provides descriptive statistics for each grade with respect to the normative sample performance, in the Enterprise scaled score units.

Table 54: Descriptive Statistics for Weighted Scaled Scores by Grade for the Norming Sample in the Unified Scale

	Fall Unified Scaled Scores				Spring Unified Scaled Scores			
Grade	**N**	**Mean**	**Standard Deviation**	**Median**	**N**	**Mean**	**Standard Deviation**	**Median**
K	219,980	702	65	702	204,950	795	69	794
1	373,300	774	76	766	247,430	856	74	856
2	563,280	886	74	888	325,560	939	70	944
3	566,780	952	69	958	351,450	987	68	990
4	624,810	995	66	1,000	341,290	1,021	67	1,024
5	502,860	1,033	66	1,038	264,970	1,056	67	1,060
6	340,390	1,065	65	1,070	161,650	1,086	68	1,090
7	213,030	1,088	68	1,092	102,030	1,105	70	1,108
8	175,210	1,111	69	1,116	61,530	1,127	72	1,130
9	57,120	1,129	71	1,134	41,130	1,138	69	1,144
10	67,800	1,139	69	1,146	33,110	1,144	70	1,152
11	38,420	1,144	71	1,152	11,710	1,149	70	1,158
12	23,420	1,153	72	1,162	4,510	1,156	72	1,164

2. As part of the development of the Star Early Learning Unified scale, Star Early Literacy Rasch scores were equated to the Star Reading Rasch scale. This resulted in a downward extension of the latter scale that encompasses the full range of both Star Early Literacy and Reading performance. This extended Rasch scale was employed to put all students' scores on the same scale for purposes of norms development.

Table 55: Descriptive Statistics for Weighted Scaled Scores by Grade for the Norming Sample in the Enterprise Scale

	Fall Enterprise Scaled Scores				Spring Enterprise Scaled Scores			
Grade	N	Mean	Standard Deviation	Median	N	Mean	Standard Deviation	Median
K	219,980	54	36	56	204,950	110	75	84
1	373,300	99	75	74	247,430	187	118	162
2	563,280	237	132	222	325,560	344	153	334
3	566,780	373	159	364	351,450	462	182	450
4	624,810	483	184	470	341,290	569	217	538
5	502,860	608	225	578	264,970	700	260	656
6	340,390	738	262	698	161,650	831	289	818
7	213,030	842	293	830	102,030	912	301	906
8	175,210	940	300	938	61,530	1,003	299	1,036
9	57,120	1,016	295	1,054	41,130	1,057	282	1,146
10	67,800	1,058	282	1,148	33,110	1,079	281	1,184
11	38,420	1,078	285	1,190	11,710	1,099	278	1,216
12	23,420	1,113	277	1,234	4,510	1,122	274	1,246

Growth Norms

Student achievement typically is thought of in terms of status: a student's performance at one point in time. However, this ignores important information about a student's learning trajectory—how much students are growing over a period of time. When educators are able to consider growth information—the amount or rate of change over time—alongside current status, a richer picture of the student emerges, empowering educators to make better instructional decisions.

To facilitate deeper understanding of achievement, Renaissance Learning maintains growth norms for Star Assessments that provide insight both on growth to date and likely growth in the future. Growth norms are currently available for Star Math, Star Reading, and Star Early Literacy, and may be available for additional Star adaptive assessments in the coming years.

The growth model used by Star Assessments is Student Growth Percentile (Betebenner, 2009). SGPs were developed by Dr. Damian Betebenner, originally in partnership with several state departments of education.[3] It should be noted that the initial development of SGP involved annual state summative tests with

3. Core SGP documentation and source code are publicly available at https://cran.r-project.org/web/packages/SGP/index.html.

reasonably constrained testing periods within each state. Because Star tests may be taken at multiple times throughout the year, a number of adaptations to the original model were made. For more information about Star Reading SGPs, please refer to this overview: http://doc.renlearn.com/KMNet/R00571375CF86BBF.pdf

SGPs are norm-referenced estimates that compare a student's growth to that of his or her academic peers nationwide. Academic peers are defined as those students in the same grade with a similar score history. SGPs are generated via a process that uses quantile regression to provide a measure of how much a student changed from one Star testing window to the next relative to other students with similar score histories. SGPs range from 1–99 and are interpreted similarly to Percentile Ranks, with 50 indicating typical or expected growth. For instance, an SGP score of 37 means that a student grew as much or more than 37 percent of her academic peers.

The Star Reading SGP package also produces a range of future growth estimates. Those are mostly hidden from users but are presented in goal setting and related applications to help users understand what typical or expected growth looks like for a given student. They are particularly useful for setting future goals and understanding the likelihood of reaching future benchmarks, such as likely achievement of proficient on an upcoming state summative assessment.

At present, the Star Reading SGP growth norms are based on a sample of 23, 376,700 matched student records from the 2016–2017, 2017–2018, and 2018–2019 school years across grades 1–12. The sample included 9,778,703 unique students across all three school years. Table 56 below provides a summary of the number of students and tests that were used when computing the SGP growth norms.

Table 56: Number of Students and Number of Tests Used in Computing SGP Growth Norms

Grade	Students	Tests	Grade	Students	Tests
1	761,260	1,691,537	8	875,682	2,035,964
2	1,142,750	2,903,543	9	466,481	994,943
3	1,234,147	3,132,872	10	380,663	792,217
4	1,247,517	3,137,316	11	261,521	512,893
5	1,260,547	3,127,739	12	180,836	322,067
6	1,060,365	2,574,598	Total	9,778,703	23,376,700
7	906,934	2,151,011			

Score Definitions

This chapter enumerates all of the scores reported by Star Reading, including scaled scores, norm-referenced, and criterion-referenced scores.

Types of Test Scores

In a broad sense, Star Reading software provides three broad types of test scores that measure student performance in different ways:

- *Criterion-referenced scores* describe a student's performance relative to a specific content domain or to a standard. Such scores may be expressed either on a continuous score scale or as a classification. An example of a criterion-referenced score on a continuous scale is a percent-correct score, which expresses what proportion of test questions the student can answer correctly in the content domain. An example of a criterion-referenced classification is a proficiency category on a standards-based assessment: the student may be said to be "proficient" or not, depending on whether the student's score equals, exceeds, or falls below a specific criterion (the "standard") used to define "proficiency" on the standards-based test. The criterion-referenced score reported by Star Reading is the Instructional Reading Level, which compares a student's test performance to the 1995 updated vocabulary lists that are based on the EDL's Core Vocabulary list. The Instructional Reading Level is the highest grade level at which the student is estimated to comprehend 80 percent of the text written at that level.
- *Norm-referenced scores* compare a student's test results to the results of other students who have taken the same test. In this case, scores provide a relative measure of student achievement compared to the performance of a group of students at a given time. Percentile Ranks and Grade Equivalents are the two primary norm-referenced scores available in Star Reading software. Both of these scores are based on a comparison of a student's test results to the data collected during the 2017 national norming program.
- *Scaled scores* are the fundamental scores used to summarize students' performance on Star Reading tests. Upon completion of the test, the testing software calculates a single-valued Star Reading Unified scale score or Star Reading Enterprise scale score. The Unified scale score is a linear transformation of the Rasch estimate, while the Enterprise scale is a non-linear transformation of the Rasch ability estimate as described below.

Enterprise Scale Scores

For Star Reading, the "Enterprise" scale scores are the same scores that have been reported continuously since Star Reading Version 1 was introduced, in 1996. Because Version 1 was not based on item response theory, its scores were expressed on an ad hoc vertical (developmental) scale related to the student's reading grade level; scale scores ranged from 50 to 1350. The use of item response theory was introduced into Star Reading Version 2. Beginning with that version, Star software calculated students' scores on the Rasch IRT ability scale. To maintain continuity with the non-IRT score scale used in Version 1, the Rasch ability scores were converted to scores on the original scale by means of an equipercentile equating transformation. At that time, the range of reported Enterprise scale scores was extended to 0 to 1400.

Unified Scale Scores

Many users of Star Reading use Star Early Literacy to assess their students until they are ready to take Star Reading itself. Until recently, Star Reading and Star Early Literacy used different score scales, making it difficult to monitor growth as students transitioned from one assessment to the other. To ameliorate that disparity in the two tests' score scales, Renaissance developed a single score scale that applies to both assessments: the Unified score scale. That development began with equating the two tests' underlying Rasch ability scales; the result was the "unified Rasch scale", which is a downward extension of the Rasch scale used in all Star Reading versions since the introduction of version 2. The next step was to develop an integer scale based on the unified Rasch scale, with scale scores anchored to important points on the original Enterprise score scales of both tests. The end result was a reported score scale that extends from 200 to 1400: Star Early Literacy Unified scale scores range from 200 to 1100; Star Reading Unified scale scores range from 600 to 1400. An added benefit of the Unified scale is an improvement in certain properties of the scale scores: Scores on both tests are much less variable from grade to grade; measurement error is likewise less variable; and Unified score reliability is slightly higher than that of the Enterprise scores.

Grade Equivalent (GE)

A Grade Equivalent (GE) indicates the grade placement of students for whom a particular score is typical. If a student receives a GE of 10.7, this means that the student scored as well on Star Reading as did the typical student in the seventh month of grade 10. It does not necessarily mean that the student can read independently at a tenth-grade level, only that he or she obtained a Scaled Score as high as the average tenth-grade, seventh-month student in the norms group.

GE scores are often misinterpreted as though they convey information about what a student knows or can do—that is, as if they were criterion-referenced scores. To the contrary, GE scores are norm-referenced.

Star Reading Grade Equivalents range from 0.0–12.9+. The scale divides the academic year into 10 monthly increments, and is expressed as a decimal with the unit denoting the grade level and the individual "months" in tenths. Table 57 indicates how the GE scale corresponds to the various calendar months. For example, if a student obtained a GE of 4.6 on a Star Reading assessment, this would suggest that the student was performing similarly to the average student in the fourth grade at the sixth month (March) of the academic year. Because Star Reading norms are based on fall and spring score data only, monthly GE scores are derived through interpolation by fitting a curve to the grade-by-grade medians. Table 58 on page 120 contains the Scaled Score to GE conversions.

Table 57: Incremental Grade Placements per Month

Month	Decimal Increment	Month	Decimal Increment
July	0.00 or 0.99[a]	January	0.4
August	0.00 or 0.99[a]	February	0.5
September	0.0	March	0.6
October	0.1	April	0.7
November	0.2	May	0.8
December	0.3	June	0.9

a. Depends on the current school year set in Renaissance.

The Grade Equivalent scale is not an equal-interval scale. For example, an increase of 50 Scaled Score points might represent only two or three months of GE change at the lower grades, but over a year of GE change in the high school grades. This is because student growth in reading (and other academic areas) is not linear; it occurs much more rapidly in the lower grades and slows greatly after the middle years. Consideration of this should be made when averaging GE scores, especially if it is done across two or more grades.

Comparing the Star Reading Test with Conventional Tests

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

Star Reading Grade Equivalent scores are more consistently accurate—even as a student's achievement level deviates from the level of grade placement. A student may be tested on any level of material, depending upon his or her actual performance on the test; students are tested on items of an appropriate level of difficulty, based on their individual level of achievement. Thus, a GE score of 7.6 indicates that the student's score can be appropriately compared to that of a typical seventh grader in the sixth month of the school year (with the same caveat as before—it does not mean that the student can actually handle seventh-grade reading material).

Estimated Oral Reading Fluency (Est. ORF)

Estimated Oral Reading Fluency (Est. ORF) is an estimate of a student's ability to read words quickly and accurately in order to comprehend text efficiently. Students with oral reading fluency demonstrate accurate decoding, automatic word recognition, and appropriate use of the rhythmic aspects of language (e.g., intonation, phrasing, pitch, and emphasis).

Est. ORF is reported as an estimated number of words a student can read correctly within a one-minute time span on grade-level-appropriate text. Grade-level text was defined to be connected text in a comprehensible passage form that has a readability level within the range of the first half of the school year. For instance, an Est. ORF score of 60 for a second-grade student would be interpreted as meaning the student is expected to read 60 words correctly within one minute on a passage with a reading grade level between 2.0 and 2.5. Therefore, when this estimate is compared to an observed score on a specific passage which has a fixed level of readability, there might be noticeable differences as the Est. ORF provides an estimate across a range of readability levels.

The Est. ORF score was computed using the results of a large-scale research study investigating the linkage between the Star Reading scores and estimates of oral reading fluency on a range of passages with grade-level-appropriate difficulty. An equipercentile linking was done between Star Reading scores and oral reading fluency, providing an estimate of the oral reading fluency for each Scaled Score unit in Star Reading for grades 1–4 independently. Results of the analysis can be found in "Additional Validation Evidence for Star Reading" on page 74. A table of selected Star Reading Scaled Scores and corresponding Estimated ORF values can be found in "Appendix B: Detailed Evidence of Star Reading Validity" on page 144.

Instructional Reading Level (IRL)

The Instructional Reading Level is a criterion-referenced score that indicates the highest reading level at which the student can effectively be taught. In other words, IRLs tell you the reading level at which students can recognize words and comprehend written instructional material with some assistance. A sixth-grade student with an IRL of 4.0, for example, would be best served by instructional materials prepared at the fourth-grade level. IRLs are represented by either numbers or letters indicating a particular grade. Number codes represent IRLs for grades 1.0–12.9. IRL letter codes include PP (Pre-Primer), P (Primer, grades .1–.9), and PHS (Post-High School, grades 13.0+).

As a construct, instructional reading levels have existed in the field of reading education for over seventy years. During this time, a variety of assessment instruments have been developed using different measurement criteria that teachers can use to estimate IRL. Star Reading software determines IRL scores relative to 1995 updated vocabulary lists that are based on the Educational Development Laboratory's (EDL) *A Revised Core Vocabulary* (1969). The Instructional Reading Level is defined as the highest reading level at which the student can read at 90–98 percent word recognition (Gickling & Haverape, 1981; Johnson, Kress & Pikulski, 1987; McCormick, 1999) and with 80 percent comprehension or higher (Gickling & Thompson, 2001). Although Star Reading does not directly assess word recognition, Star Reading uses the student's Rasch ability scores, in conjunction with the Rasch difficulty parameters of graded vocabulary items, to determine the proportion of items a student can comprehend at each grade level.

Special IRL Scores

If a student's performance on Star Reading indicates an IRL below the first grade, Star Reading software will automatically assign an IRL score of Primer (P) or Pre-Primer (PP). Because the kindergarten-level test items are designed so that even readers of very early levels can understand them, a Primer or Pre-Primer IRL means that the student is essentially a non-reader. There are, however, other unusual circumstances that could cause a student to receive an IRL of Primer or Pre-Primer. Most often, this happens when a student simply does not try or purposely answers questions incorrectly.

When Star Reading software determines that a student can answer 80 percent or more of the grade 13 items in the Star Reading test correctly, the student is assigned an IRL of Post-High School (PHS). This is the highest IRL that anyone can obtain when taking the Star Reading test.

Understanding IRL and GE Scores

One strength of Star Reading software is that it provides both criterion-referenced and norm-referenced scores. As such, it provides more than one frame of reference for describing a student's current reading performance. The two frames of reference differ significantly, however, so it is important to understand the two estimates and their development when making interpretations of Star Reading results.

The Instructional Reading Level (IRL) is a criterion-referenced score. It provides an estimate of the grade level of written material with which the student can most effectively be taught. While the IRL, like any test result, is simply an estimate, it provides a useful indication of the level of material on which the student should be receiving instruction. For example, if a student (regardless of current grade placement) receives a Star Reading IRL of 4.0, this indicates that the student can most likely learn without experiencing too many difficulties when using materials written to be on a fourth-grade level.

The IRL is estimated based on the student's pattern of responses to the Star Reading items. A given student's IRL is the highest grade level of items at which it is estimated that the student can correctly answer at least 80 percent of the items.

In effect, the IRL references each student's Star Reading performance to the difficulty of written material appropriate for instruction. This is a valuable piece of information in planning the instructional program for individuals or groups of students.

The Grade Equivalent (GE) is a norm-referenced score. It provides a comparison of a student's performance with that of other students around the nation. If a student receives a GE of 4.0, this means that the student scored as well on the Star Reading test as did the typical student at the beginning of grade 4. It does not mean that the student can read books that are written at a fourth-grade level—only that he or she reads as well as fourth-grade students in the norms group.

In general, IRLs and GEs will differ. These differences are caused by the fact that the two score metrics are designed to provide different information. That is, IRLs estimate the level of text that a student can read with some instructional assistance; GEs express a student's performance in terms of the grade level for which that performance is typical. Usually, a student's GE score will be higher than the IRL.

The score to be used depends on the information desired. If a teacher or educator wishes to know how a student's Star Reading score compares with that of other students across the nation, either the GE or the Percentile Rank should be used. If the teacher or educator wants to know what level of instructional materials a student should be using for ongoing classroom schooling, the IRL is the preferred score. Again, both scores are estimates of a student's current level of reading achievement. They simply provide two ways of interpreting this performance—relative to a national sample of students (GE) or relative to the level of written material the student can read successfully (IRL).

Percentile Rank (PR)

Percentile Rank is a norm-referenced score that indicates the percentage of students in the same grade and at the same point of time in the school year who obtained scores lower than the score of a particular student. In other words, Percentile Ranks show how an individual student's performance compares to that of his or her same-grade peers on the national level. For example, a Percentile Rank of 85 means that the student is performing at a level that exceeds 85 percent of other students in that grade at the same time of the year. Percentile Ranks simply indicate how a student performed compared to the others who took Star Reading tests as a part of the national norming program. The range of Percentile Ranks is 1–99.

The Percentile Rank scale is not an equal-interval scale. For example, for a student with a grade placement of 1.7, a Unified Scaled Score of 896 corresponds to a PR of 80, and a Unified Scaled Score of 931 corresponds to a PR of 90. Thus, a difference of 35 Scaled Score points represents a 10-point difference in PR. However, for students at the same 1.7 grade placement, a Unified Scaled Score of 836 corresponds to a PR of 50, and a Unified Scaled Score of 853 corresponds to a PR of 60. While there is now only a 17-point difference in Scaled Scores, there is still a 10-point difference in PR. For this reason, PR scores should not be averaged or otherwise algebraically manipulated. NCE scores are much more appropriate for these activities.

Table 59 on page 124 and Table 60 on page 127 contain an abridged version of the Scaled Score to Percentile Rank conversion table that the Star Reading software uses. The actual table includes data for all of the monthly grade placement values from 0.0–12.9.

This table can be used to estimate PR values for tests that were taken when the grade placement value of a student was incorrect (see "Types of Test Scores" on page 106 for more information). If the error is caught right away, one always has the option of correcting the grade placement for the student and then having the student retest.

Normal Curve Equivalent (NCE)

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

Table 61 on page 131 provides the NCEs corresponding to integer PR values and facilitates the conversion of PRs to NCEs. Table 62 on page 132 provides the conversions from NCE to PR. The NCE values are given as a range of scores that convert to the corresponding PR value.

Student Growth Percentile (SGP)

Student Growth Percentiles (SGPs) are a norm-referenced quantification of individual student growth derived using quantile regression techniques. An SGP compares a student's growth to that of his or her academic peers nationwide with a similar achievement history on Star assessments. Academic peers are students who

- are in the same grade,
- had the same scores on the current test and (up to) two prior tests from different windows of testing time, and
- took the most recent test and the first prior test on the same dates.

SGPs provide a measure of how a student changed from one Star testing window[1] to the next relative to other students with similar starting Star Reading scores. SGPs range from 1–99 and interpretation is similar to that of Percentile Rank scores; lower numbers indicate lower relative growth and higher numbers show higher relative growth. For example, an SGP of 70 means that the student's growth from one test window to another exceeds the growth of 70% of students nationwide in the same grade with a similar Star Reading score history. All students, no matter their starting Star score, have an equal chance to demonstrate growth at any of the 99 percentiles.

SGPs are often used to indicate whether a student's growth is more or less than can be expected. For example, without an SGP, a teacher would not know if a Scaled Score increase of 100 points represents good, not-so-good, or average

1. We collect data for our growth norms during three different time periods: fall, winter, and spring. More information about these time periods is provided on page 114.

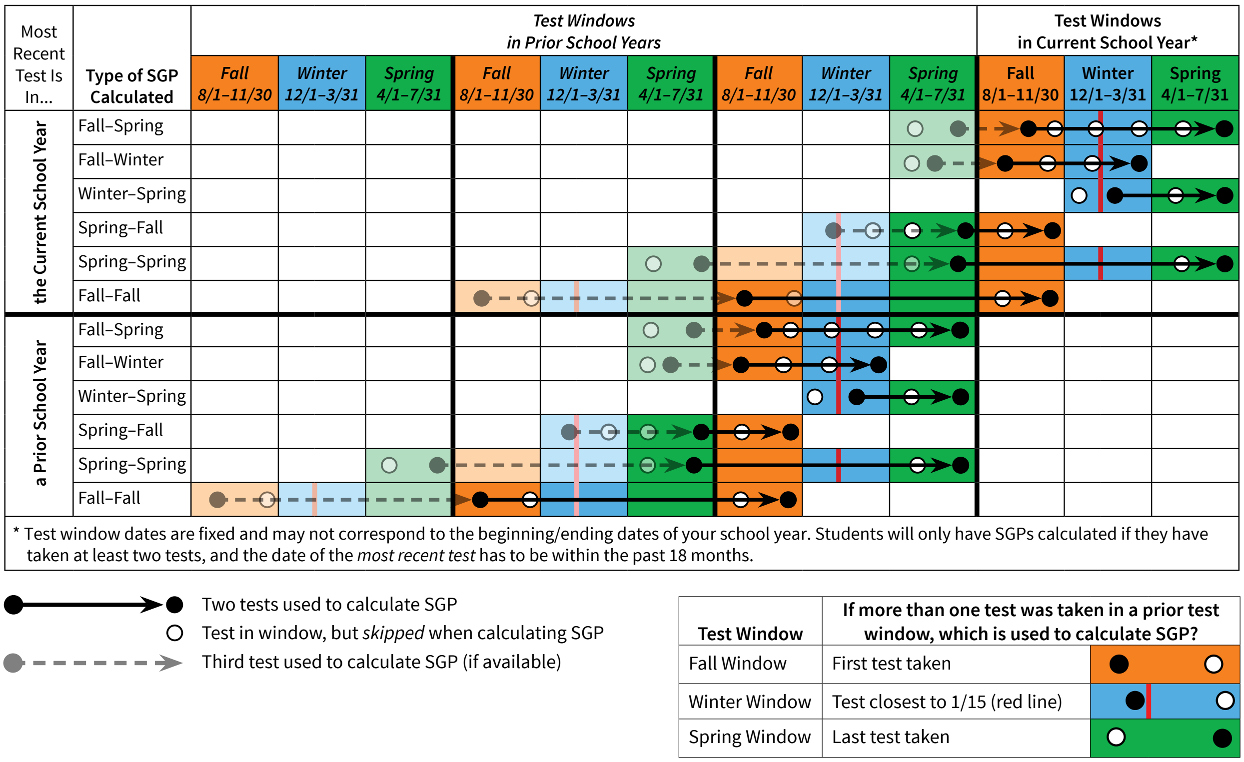
growth. This is because students of differing achievement levels in different grades grow at different rates relative to the Star Reading scale. For example, a high-achieving second-grader grows at a different rate than a low-achieving second-grader. Similarly, a high-achieving second-grader grows at a different rate than a high-achieving eighth-grader.

SGPs can be aggregated to describe typical growth for groups of students—for example, a class, grade, or school as a whole—by calculating the group's median, or middle, growth percentile. No matter how SGPs are aggregated, whether at the class, grade, or school level, the statistic and its interpretation remain the same. For example, if the students in one class have a median SGP of 62, that particular group of students, on average, achieved higher growth than their academic peers.

SGP is calculated for students who have taken at least two tests (a *current* test and a *prior* test) within at least two different testing windows (Fall, Winter, or Spring).

If a student has taken more than one test in a single test window, the SGP calculation is based off the following tests:

- The current test is always the last test taken in a testing window.
- The test used as the prior test depends on what testing window it falls in:
    - Fall window: The first test taken in the Fall window is used.
    - Winter window: The test taken closest to January 15 in the Winter window is used.
    - Spring window: The last test taken in the Spring window is used.



Lexile® Measures

In cooperation with MetaMetrics®, since August 2014, users of Star Reading have had the option of including Lexile measures on certain Star Reading score reports. Reported Lexile measures will range from BR400L to 1825L. (The “L” suffix identified the score as a Lexile measure. Where it appears, the “BR” prefix indicates a score that is below 0 on the Lexile scale; such scores are typical of beginning readers.)

Lexile Measures of Students and Books: Measures of Student Reading Achievement and Text Readability

The ability to read and comprehend written text is important for academic success. Students may, however, benefit most from reading materials that match their reading ability/achievement: reading materials that are neither too easy nor too hard so as to maximize learning. To facilitate students’ choices of appropriate reading materials, measures commonly referred to as *readability measures* are used in conjunction with students’ reading achievement measures.

A text readability measure can be defined as a numeric scale, often derived analytically, that takes into account text characteristics that influence text comprehension or readability. An example of a readability measure is an age-level estimate of text difficulty. Among text characteristics that can affect text comprehension are sentence length and word difficulty.

A person's reading measure is a numeric score obtained from a reading achievement test, usually a standardized test such as Star Reading. A person's reading score quantifies his/her reading achievement level at a particular point in time.

Matching a student with text/books that target a student's interest and level of reading achievement is a two-step process: first, a student's reading achievement score is obtained by administering a standardized reading achievement test; second, the reading achievement score serves as an entry point into the readability measure to determine the difficulty level of text/books that would best support independent reading for the student. Optimally, a readability measure should match students with books that they are able to read and comprehend independently without boredom or frustration: books that are engaging yet slightly challenging to students based on the students' reading achievement and grade level.

Renaissance Learning's (RLI) readability measure is known as the Advantage/TASA Open Standard for Readability (ATOS). The *ATOS for Text* readability formula was developed through extensive research by RLI in conjunction with Touchstone Applied Science Associates, Inc. (TASA), now called Questar Assessment, Inc. A great many school libraries use ATOS book levels to index readability of their books. ATOS book levels, which are derived from *ATOS for Books* measures, express readability as grade levels; for example, an ATOS readability measure of 4.2 means that the book is at a difficulty level appropriate for students reading at a level typical of students in the 4th grade, 2nd month. To match students to books at an appropriate level, the widely used Accelerated Reader system uses ATOS measures of readability and student's Grade Equivalent (GE) scores on standardized reading tests such as Star Reading.

Another widely-used system for matching readers to books at appropriate difficulty levels is The Lexile Framework® for Reading, developed by MetaMetrics, Inc. The Lexile scale is a common scale for both text measure (readability or text difficulty) and reader measure (reading achievement scores); in the Lexile Framework, both text difficulty and person reading ability are measured on the same scale. Unlike *ATOS for Books,* the Lexile Framework expresses a book's reading difficulty level (and students' reading ability levels) on a continuous scale ranging from below 0 to 1825 or more. Because some schools and school libraries use the Lexile Framework to index the reading difficulty levels of their books, there was a need to provide users of Star Reading with a student reading ability score compatible with the Lexile Framework.

In 2014, Metametrics, Inc., developed a means to translate Star Reading scale scores into equivalent Lexile measures of student reading ability. To do so, more than 200 MetaMetrics reading test items that had already been calibrated on the Lexile scale were administered in small numbers as unscored scale anchor items at the end of Star Reading tests. More than 250,000 students in grades 1 through 12 took up to 6 of those items as part of their Star Reading tests in April 2014. MetaMetrics' analysis of the Star Reading and Lexile anchor item response data yielded a means of transforming Star Reading's underlying Rasch scores into equivalent Lexile scores. That transformation, in turn, was used to develop a concordance table listing the Lexile equivalent of each unique Star Reading scale score.

In some cases, a range of text/book reading difficulty in which a student can read independently or with minimal guidance is desired. At Renaissance, we define the range of reading difficulty level that is neither too hard nor too easy as the Zone of Proximal Development (ZPD). The ZPD range allows, potentially, optimal learning to occur because students are engaged and appropriately challenged by reading materials that match their reading achievement and interest. The ZPD range is simply an approximation of the range of reading materials that is likely to benefit the student most. ZPD ranges are not absolute and teachers should also use their objective judgment to help students select reading books that enhance learning.

In a separate linking procedure, MetaMetrics compared the ATOS readability measures of thousands of books to the Lexile measures of the same books. Analysis of those data yielded a table of equivalence between ATOS reading grade levels and Lexile readability measures. That equivalence table supports matching students to books regardless of whether a book's readability is measured using the Renaissance Learning ATOS system or the Lexile Framework created by MetaMetrics. Additionally, it supports translating ATOS ZPD ranges into equivalent ZPD ranges expressed on the Lexile scale.

Special Star Reading Scores

Most of the scores provided by Star Reading software are common measures of reading performance. Star Reading software also determines the Zone of Proximal Development.

Zone of Proximal Development (ZPD)

The Zone of Proximal Development (ZPD) defines the readability range from which students should be selecting books in order to ensure sufficient comprehension and therefore achieve optimal growth in reading skills without experiencing frustration. Star Reading software uses Grade Equivalents to derive a student's

ZPD score. Specifically, it relates the Grade Equivalent estimate of a student's reading ability with the range of most appropriate readability levels to use for reading practice. Table 63 on page 133 shows the relationship between GEs and ZPD scores.

The Zone of Proximal Development is especially useful for students who use Accelerated Reader, which provides readability levels on over 180,000 trade books and textbooks. Renaissance Learning developed the ZPD ranges according to Vygotskian theory, based on an analysis of Accelerated Reader book reading data from 80,000 students in the 1996–1997 school year. More information is available in *The research foundation for Accelerated Reader goal-setting practices* (2006), which is published by Renaissance Learning (http://doc.renlearn.com/KMNet/R001438603GC81D6.pdf).

Grade Placement

Star Reading software uses the student's grade placement—grade and month of the school year—when determining the norm-referenced scores. The values of PR and NCE are based not only on what scaled score the student achieved but also on the grade placement of the student at the time of the test (for example, a second-grader in the seventh month with a Unified scaled score of 957 would have a PR of 34, while a third-grader in the seventh month with the same scaled score would have a PR of 41). Thus, it is crucial that student records indicate the proper grade when students take a Star Reading test, and that any testing in July or August reflects the proper understanding of how Star Reading software deals with these months in determining grade placement.

Indicating the Appropriate Grade Placement

The numeric representation of a student's grade placement is based on the specific month and day in which he or she takes a test. Although teachers indicate a student's grade level using whole numbers, Star Reading software automatically adds fractional increments to that grade level based on the month and day of the test. (Note: Grade placements for pre-K students are negative numbers.)
To determine the appropriate increment, Star Reading software considers the standard school year to run from September—June and assigns increment values of .0–.9 to these months. Table 57 on page 108 summarizes the increment values assigned to each month.

The increment values for July and August depend on the school year setting:

- If teachers will use the July and August test scores to evaluate the student's reading performance at the beginning of the year, educators must make sure the following school year is set as the current school year in the Renaissance program at the time they administer the summer tests. Grades are automatically increased by one level in each successive school year, so promoting students to the next grade is not necessary. In this case, the increment value for July and August is 0.00 because these months are at the beginning of the school year.

- If teachers will use the test scores to evaluate the student's reading performance at the end of the school year, they must make sure the school year that has just ended is set as the current school year in the Renaissance program at the time they administer the summer tests. In this case, the increment value for July and August is 0.99 because these months are at the end of the school year that has passed.

In addition to the tenths digit appended to the grade level to denote the month of the standard school year in which a test was taken, Star Reading appends a hundredths digit to denote the day on which a test was taken as well. The hundredths digit represents the fractional portion of a 30-day month. For example, the increment for a test taken on the sixth day of the month is 0.02. For a test taken on the twenty-fourth day of the month, the increment is 0.08.

If a school follows the standard school calendar used in Star Reading software and does not test in the summer, assigning the appropriate grade placements for students is relatively easy. However, if students will be tested in July or August—whether it is for a summer reading program or because the normal calendar extends into these months—grade placements become an extremely important issue.

To ensure the accurate determination of norm-referenced scores when testing in the summer, it must be determined when to set the next school year as the current school year, and thereby advance students from one grade to the next. In most cases, the guidelines above can be used.

Instructions for specifying school years and grade assignments can be found at https://help.renaissance.com/setup.

Compensating for Incorrect Grade Placements

Teachers cannot make retroactive corrections to a student's grade placement by editing the grade assignments in a student's record or by adjusting the increments for the summer months after students have tested. In other words, Star Reading software cannot go back in time and correct scores resulting from erroneous grade placement information. Thus, it is extremely important for the test administrator to make sure that the proper grade placement procedures are being followed.

Conversion Tables

Table 58: Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
0.0	600	700	0	55
0.1	701	708	56	57
0.2	709	716	58	60
0.3	717	725	61	62
0.4	726	733	63	64
0.5	734	742	65	67
0.6	743	750	68	69
0.7	751	759	70	72
0.8	760	768	73	74
0.9	769	777	75	77
1.0	778	786	78	80
1.1	787	795	81	84
1.2	796	804	85	87
1.3	805	813	88	91
1.4	814	821	92	98
1.5	822	830	99	106
1.6	831	839	107	125
1.7	840	847	126	142
1.8	848	855	143	158
1.9	856	864	159	175
2.0	865	872	176	190
2.1	873	880	191	206
2.2	881	887	207	219
2.3	888	895	220	235
2.4	896	902	236	248
2.5	903	909	249	262
2.6	910	916	263	276
2.7	917	923	277	289
2.8	924	930	290	305
2.9	931	936	306	318
3.0	937	943	319	333
3.1	944	949	334	346

Table 58: Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
3.2	950	955	347	361
3.3	956	961	362	371
3.4	962	966	372	382
3.5	967	972	383	398
3.6	973	977	399	411
3.7	978	982	412	425
3.8	983	987	426	440
3.9	988	992	441	452
4.0	993	996	453	460
4.1	997	1001	461	471
4.2	1002	1005	472	481
4.3	1006	1009	482	493
4.4	1010	1013	494	504
4.5	1014	1017	505	516
4.6	1018	1021	517	527
4.7	1022	1024	528	536
4.8	1025	1028	537	551
4.9	1029	1032	552	560
5.0	1033	1035	561	569
5.1	1036	1038	570	580
5.2	1039	1041	581	589
5.3	1042	1044	590	600
5.4	1045	1048	601	614
5.5	1049	1050	615	621
5.6	1051	1053	622	632
5.7	1054	1056	633	643
5.8	1057	1059	644	654
5.9	1060	1062	655	667
6.0	1063	1064	668	675
6.1	1065	1067	676	688
6.2	1068	1070	689	702
6.3	1071	1072	703	711
6.4	1073	1075	712	725
6.5	1076	1077	726	737
6.6	1078	1079	738	752
6.7	1080	1082	753	774

Table 58: Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
6.8	1083	1084	775	784
6.9	1085	1087	785	799
7.0	1088	1089	800	810
7.1	1090	1091	811	823
7.2	1092	1094	824	841
7.3	1095	1096	842	850
7.4	1097	1098	851	858
7.5	1099	1100	859	870
7.6	1101	1102	871	881
7.7	1103	1105	882	894
7.8	1106	1107	895	901
7.9	1108	1109	902	907
8.0	1110	1111	908	912
8.1	1112	1113	913	920
8.2	1114	1115	921	930
8.3	1116	1117	931	941
8.4	1118	1119	942	951
8.5	1120	1121	952	963
8.6	1122	1123	964	970
8.7	1124	1125	971	980
8.8	1126	1126	981	985
8.9	1127	1128	986	1003
9.0	1129	1130	1004	1025
9.1	1131	1132	1026	1042
9.2	1133	1133	1043	1050
9.3	1134	1135	1051	1065
9.4	1136	1136	1066	1073
9.5	1137	1138	1074	1095
9.6	1139	1139	1096	1101
9.7	1140	1140	1102	1106
9.8	1141	1142	1107	1123
9.9	1143	1143	1124	1132
10.0	1144	1144	1133	1140
10.1	1145	1145	1141	1148
10.2	1146	1146	1149	1154
10.3	1147	1147	1155	1161

Table 58: Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
10.4	1148	1148	1162	1166
10.5	1149	1149	1167	1171
10.6	1150	1150	1172	1175
10.7	1151	1151	1176	1180
10.8	1152	1152	1181	1185
10.9	1153	1153	1186	1190
11.0	1154	1154	1191	1196
11.1	1155	1155	1197	1203
11.2	1156	1156	1204	1208
11.3	1157	1157	1209	1213
11.4	1158	1158	1214	1216
11.5	1159	1159	1217	1219
11.6	1160	1160	1220	1223
11.7	1161	1161	1224	1227
11.8	1162	1162	1228	1231
11.9	1163	1163	1232	1236
12.0	1164	1164	1237	1242
12.1	1165	1165	1243	1248
12.2	1166	1166	1249	1252
12.3	1167	1167	1253	1255
12.4	1168	1168	1256	1260
12.5	1169	1169	1261	1265
12.6	1170	1170	1266	1271
12.7	1171	1171	1272	1277
12.8	1172	1172	1278	1283
12.9	1173	1173	1284	1289
13.0	1174	1400	1290	1400

Table 59: Enterprise Scaled Score to Percentile Rank Conversions[a]

PR	Grade Placement												
	K	1	2	3	4	5	6	7	8	9	10	11	12
1	—	0	0	0	0	0	0	0	0	0	0	0	0
2	—	8	66	81	111	188	253	294	332	352	376	371	371
3	—	15	70	87	153	226	292	339	374	406	444	435	451
4	—	24	73	95	180	255	323	369	412	453	485	474	494
5	—	30	75	105	203	275	350	394	444	479	522	511	534
6	0	34	76	119	220	292	369	415	463	508	557	548	567
7	2	38	78	137	234	308	383	439	482	537	584	581	597
8	6	42	80	153	247	323	401	455	501	559	612	609	629
9	8	46	82	166	261	339	418	465	522	581	637	633	655
10	11	49	83	176	271	347	429	477	534	597	651	651	680
11	13	51	85	184	279	360	441	485	548	609	664	668	703
12	15	53	86	191	284	367	451	497	559	622	680	680	722
13	17	55	87	197	292	372	459	508	570	637	693	698	753
14	18	56	88	205	299	380	465	517	581	648	708	717	775
15	20	57	89	211	303	389	472	525	590	660	722	731	795
16	21	58	90	217	311	397	479	534	601	673	738	753	817
17	23	59	92	222	317	403	489	545	612	684	753	780	836
18	24	60	95	226	323	409	497	555	622	703	775	790	851
19	26	60	97	232	327	418	501	559	633	712	785	806	865
20	27	61	99	236	334	424	508	567	640	726	795	817	882
21	28	61	101	241	339	432	517	574	651	738	811	836	891
22	30	62	104	247	343	439	522	584	660	762	824	851	902
23	32	63	105	251	350	446	528	590	673	775	836	859	908
24	33	63	108	257	355	451	534	597	680	785	847	878	917
25	34	64	111	261	360	455	545	609	689	800	859	887	925
26	35	64	116	265	363	459	552	615	703	811	871	899	942
27	36	65	123	271	367	463	557	622	712	824	882	905	952
28	37	65	128	275	371	467	561	629	722	836	895	913	964
29	38	65	133	279	374	472	567	637	731	847	899	921	971
30	40	66	139	282	378	477	574	644	744	855	905	931	981
31	41	67	143	288	383	482	581	651	762	865	913	937	993
32	42	67	147	292	389	489	584	660	775	878	921	947	1015

a. Each entry is the lowest Scaled Score for that grade and percentile.

Table 59: Enterprise Scaled Score to Percentile Rank Conversions[a]

PR	Grade Placement												
	K	1	2	3	4	5	6	7	8	9	10	11	12
33	43	67	153	296	394	494	590	668	785	887	931	958	1035
34	44	68	157	301	399	499	597	673	795	895	937	964	1051
35	45	68	161	303	401	501	601	680	806	902	947	971	1066
36	46	69	164	308	406	508	609	689	817	908	958	981	1086
37	47	69	170	313	412	514	615	698	830	913	964	993	1102
38	48	69	174	317	418	519	619	708	836	921	971	1004	1115
39	48	70	178	321	421	525	626	712	847	925	981	1026	1133
40	49	70	182	325	426	528	629	722	855	937	986	1043	1149
41	50	71	185	327	432	530	637	731	859	947	1004	1058	1155
42	51	71	190	332	439	537	640	738	871	952	1026	1074	1167
43	51	71	194	337	444	545	648	753	882	964	1043	1096	1176
44	52	72	197	341	446	548	651	769	891	968	1051	1107	1181
45	53	72	201	343	451	555	660	775	899	975	1066	1124	1191
46	53	73	205	347	455	557	664	785	905	981	1086	1141	1197
47	54	73	209	352	457	559	673	795	908	993	1102	1155	1209
48	55	73	213	355	461	564	676	800	913	1015	1107	1167	1217
49	56	74	215	360	463	567	684	811	921	1026	1124	1176	1220
50	56	74	219	363	467	574	693	824	931	1043	1141	1181	1228
51	56	75	222	365	469	577	698	830	937	1058	1149	1191	1232
52	57	75	226	369	474	584	708	842	947	1074	1162	1204	1243
53	57	76	230	371	477	587	717	851	958	1096	1167	1209	1249
54	58	76	234	374	482	593	722	855	964	1102	1176	1217	1256
55	58	77	236	378	485	597	731	859	971	1115	1181	1220	1266
56	59	77	239	380	492	606	744	871	981	1133	1191	1228	1272
57	59	78	243	386	494	609	762	882	986	1149	1197	1237	1284
58	60	79	247	389	499	615	769	891	1004	1162	1209	1243	1293
59	60	79	251	394	501	619	780	899	1026	1172	1217	1253	1296
60	60	80	255	399	505	626	785	902	1035	1181	1220	1256	1298
61	61	81	259	401	511	629	795	908	1051	1186	1228	1266	1304
62	61	82	263	406	514	637	800	913	1066	1197	1232	1272	1307
63	61	82	269	412	519	640	811	921	1074	1204	1243	1284	1309

a. Each entry is the lowest Scaled Score for that grade and percentile.

Table 59: Enterprise Scaled Score to Percentile Rank Conversions[a]

PR	Grade Placement												
	K	1	2	3	4	5	6	7	8	9	10	11	12
64	62	83	273	415	525	648	817	925	1096	1214	1253	1293	1314
65	62	84	277	421	528	655	830	937	1107	1220	1261	1296	1315
66	63	85	281	426	534	660	836	947	1124	1228	1272	1301	1318
67	63	85	284	432	541	668	847	958	1141	1232	1284	1307	1319
68	64	86	288	439	545	676	855	968	1155	1249	1290	1311	1321
69	64	87	292	444	552	680	859	971	1167	1253	1296	1314	1322
70	65	88	296	446	555	689	871	981	1172	1261	1301	1316	1324
71	65	88	301	451	559	693	882	986	1181	1266	1304	1319	1327
72	65	90	306	455	564	703	891	1004	1191	1278	1307	1321	1328
73	66	90	311	459	567	712	895	1026	1204	1290	1311	1322	1329
74	67	92	317	463	574	717	902	1043	1214	1296	1315	1324	1330
75	67	95	321	467	581	726	908	1058	1220	1301	1318	1327	1332
76	67	97	325	472	584	738	913	1086	1228	1307	1319	1328	1334
77	68	100	330	477	590	753	921	1102	1237	1311	1321	1330	1336
78	68	104	337	482	597	769	931	1115	1249	1316	1323	1331	1338
79	69	106	341	489	606	780	942	1141	1256	1319	1325	1333	1340
80	69	111	345	494	609	785	947	1155	1261	1321	1327	1335	1342
81	70	121	352	499	615	795	958	1172	1272	1322	1328	1336	1342
82	71	131	357	505	622	811	968	1181	1284	1324	1330	1337	1343
83	71	137	363	511	629	824	975	1191	1293	1327	1331	1339	1343
84	72	145	369	517	640	836	986	1204	1298	1328	1333	1340	1343
85	73	153	374	525	648	847	1004	1217	1307	1330	1335	1342	1344
86	74	161	380	530	655	859	1026	1224	1311	1332	1337	1342	1345
87	75	168	389	541	668	871	1051	1237	1316	1335	1340	1343	1345
88	76	174	397	552	676	887	1066	1249	1320	1337	1342	1343	1345
89	77	182	403	559	689	902	1102	1261	1323	1340	1343	1344	1345
90	78	191	412	567	703	908	1124	1278	1327	1343	1343	1345	1346
91	80	203	424	577	717	917	1149	1290	1328	1343	1344	1345	1346
92	82	215	435	587	738	931	1167	1298	1330	1343	1345	1345	1346
93	84	226	449	597	762	952	1186	1307	1333	1344	1345	1346	1346
94	86	241	459	612	790	968	1204	1314	1335	1345	1346	1346	1347

a. Each entry is the lowest Scaled Score for that grade and percentile.

Table 59: Enterprise Scaled Score to Percentile Rank Conversions[a]

PR	Grade Placement												
	K	1	2	3	4	5	6	7	8	9	10	11	12
95	88	257	472	629	811	986	1220	1319	1339	1345	1346	1346	1347
96	91	277	489	648	851	1035	1237	1322	1342	1346	1346	1347	1347
97	103	299	508	680	887	1096	1266	1328	1344	1346	1347	1347	1347
98	135	332	541	717	921	1167	1298	1334	1345	1347	1347	1347	1348
99	190	389	587	811	986	1228	1320	1343	1346	1347	1348	1348	1349

a. Each entry is the lowest Scaled Score for that grade and percentile.

Table 60: Unified Scaled Score to Percentile Rank Conversions[a]

PR	Grade Placement												
	K	1	2	3	4	5	6	7	8	9	10	11	12
1	—	600	600	600	600	600	600	600	600	600	600	600	600
2	—	614	736	786	834	871	905	926	943	952	964	961	961
3	—	627	750	803	853	891	925	946	963	976	989	986	992
4	—	643	759	817	867	906	939	960	978	993	1007	1003	1010
5	—	654	766	828	879	916	951	971	989	1005	1020	1016	1024
6	600	662	772	837	888	925	960	979	998	1015	1031	1028	1035
7	604	669	778	845	895	932	967	987	1006	1025	1040	1039	1044
8	610	676	783	853	902	939	974	994	1013	1032	1048	1047	1053
9	615	683	788	860	909	946	980	999	1020	1039	1055	1054	1060
10	620	688	792	865	914	950	984	1004	1024	1044	1059	1059	1066
11	624	692	796	869	918	955	988	1007	1028	1047	1062	1063	1071
12	627	696	799	873	921	959	992	1011	1032	1051	1066	1066	1075
13	630	699	803	876	925	962	996	1015	1036	1055	1069	1070	1080
14	633	702	806	880	928	966	999	1018	1039	1058	1072	1074	1083
15	636	705	808	883	930	969	1002	1021	1042	1061	1075	1077	1087
16	638	708	811	886	933	972	1005	1024	1045	1064	1078	1080	1091
17	641	710	814	889	936	975	1008	1027	1048	1067	1080	1084	1094
18	644	713	817	891	939	977	1011	1030	1051	1071	1083	1086	1097
19	647	715	819	894	941	980	1013	1032	1054	1073	1085	1089	1100
20	649	717	822	896	944	982	1015	1035	1056	1076	1087	1091	1103

a. Each entry is the lowest Scaled Score for that grade and percentile.

Table 60: Unified Scaled Score to Percentile Rank Conversions[a]

PR	Grade Placement												
	K	1	2	3	4	5	6	7	8	9	10	11	12
21	651	719	824	899	946	985	1018	1037	1059	1078	1090	1094	1105
22	654	722	827	902	948	987	1020	1040	1061	1081	1092	1097	1108
23	657	724	829	904	951	990	1022	1042	1064	1083	1094	1099	1110
24	659	726	832	907	953	992	1024	1044	1066	1085	1096	1102	1113
25	661	728	834	909	955	994	1027	1047	1068	1088	1099	1104	1115
26	663	730	836	911	957	996	1029	1049	1071	1090	1101	1107	1118
27	665	731	839	914	959	998	1031	1051	1073	1092	1103	1109	1120
28	666	733	841	916	961	1000	1033	1053	1075	1094	1106	1112	1122
29	669	734	843	918	963	1002	1035	1055	1077	1096	1107	1114	1124
30	671	736	846	920	965	1004	1037	1057	1079	1098	1109	1116	1126
31	673	738	848	923	967	1006	1039	1059	1081	1100	1112	1117	1128
32	676	739	850	925	969	1008	1040	1061	1083	1102	1114	1119	1130
33	678	740	853	927	971	1010	1042	1063	1085	1104	1116	1121	1132
34	679	742	855	929	973	1012	1044	1064	1087	1106	1117	1122	1134
35	681	743	857	930	974	1013	1045	1066	1089	1108	1119	1124	1136
36	683	745	859	932	976	1015	1047	1068	1091	1110	1121	1126	1138
37	684	746	862	934	978	1017	1049	1070	1093	1112	1122	1128	1140
38	686	748	864	936	980	1019	1050	1072	1094	1114	1124	1129	1142
39	687	749	866	938	981	1021	1052	1073	1096	1115	1126	1131	1144
40	688	750	868	940	983	1022	1053	1075	1098	1117	1127	1133	1146
41	690	752	870	941	985	1023	1055	1077	1099	1119	1129	1135	1147
42	691	753	872	943	987	1025	1056	1078	1101	1120	1131	1137	1149
43	692	755	874	945	989	1027	1058	1080	1103	1122	1133	1139	1151
44	693	756	876	947	990	1028	1059	1082	1105	1123	1134	1141	1152
45	695	757	878	948	992	1030	1061	1083	1107	1125	1136	1143	1154
46	696	759	880	950	994	1031	1062	1085	1109	1126	1138	1145	1155
47	697	760	882	952	995	1032	1064	1087	1110	1128	1140	1147	1157
48	699	762	884	953	997	1034	1065	1088	1112	1130	1141	1149	1159
49	700	763	885	955	998	1035	1067	1090	1114	1131	1143	1151	1160
50	701	765	887	957	1000	1037	1069	1092	1116	1133	1145	1152	1162
51	702	767	889	958	1001	1038	1070	1093	1117	1135	1146	1154	1163
52	704	768	891	960	1003	1040	1072	1095	1119	1137	1148	1156	1165

a. Each entry is the lowest Scaled Score for that grade and percentile.

Table 60: Unified Scaled Score to Percentile Rank Conversions[a]

PR	Grade Placement												
	K	1	2	3	4	5	6	7	8	9	10	11	12
53	705	770	893	961	1004	1041	1074	1097	1121	1139	1149	1157	1166
54	707	772	895	963	1006	1043	1075	1098	1122	1140	1151	1159	1168
55	708	774	896	965	1007	1044	1077	1099	1124	1142	1152	1160	1170
56	710	776	898	966	1009	1046	1079	1101	1126	1144	1154	1162	1171
57	711	778	900	968	1010	1047	1081	1103	1127	1146	1155	1164	1173
58	713	780	902	969	1012	1049	1082	1105	1129	1148	1157	1165	1175
59	714	782	904	971	1013	1050	1084	1107	1131	1150	1159	1167	1176
60	715	784	906	973	1014	1052	1085	1108	1132	1152	1160	1168	1177
61	717	786	908	974	1016	1053	1087	1110	1134	1153	1162	1170	1179
62	718	788	910	976	1017	1055	1088	1112	1136	1155	1163	1171	1180
63	720	790	913	978	1019	1056	1090	1114	1137	1156	1165	1173	1181
64	721	792	915	979	1021	1058	1091	1115	1139	1158	1167	1175	1183
65	723	794	917	981	1022	1060	1093	1117	1141	1160	1169	1176	1184
66	725	796	919	983	1024	1061	1094	1119	1143	1162	1171	1178	1186
67	726	798	921	985	1026	1063	1096	1121	1145	1163	1173	1180	1187
68	728	800	923	987	1027	1065	1098	1123	1147	1166	1174	1182	1189
69	729	802	925	989	1029	1066	1099	1124	1149	1167	1176	1183	1190
70	731	804	927	990	1030	1068	1101	1126	1150	1169	1178	1185	1192
71	733	806	929	992	1032	1069	1103	1127	1152	1170	1179	1187	1194
72	734	809	931	994	1034	1071	1105	1129	1154	1172	1180	1189	1196
73	736	811	933	996	1035	1073	1106	1131	1156	1174	1182	1190	1197
74	738	814	936	998	1037	1074	1108	1133	1158	1176	1184	1192	1199
75	739	817	938	1000	1039	1076	1110	1135	1160	1178	1186	1194	1201
76	741	820	940	1002	1040	1078	1112	1138	1162	1180	1187	1196	1203
77	743	823	942	1004	1042	1080	1114	1140	1164	1182	1189	1198	1205
78	744	827	945	1006	1044	1082	1116	1142	1166	1185	1191	1200	1207
79	746	830	947	1008	1046	1084	1118	1145	1168	1187	1193	1202	1209
80	748	834	949	1010	1047	1085	1119	1147	1169	1189	1194	1204	1211
81	750	838	952	1012	1049	1087	1121	1150	1171	1190	1196	1205	1212
82	753	842	954	1014	1051	1090	1123	1152	1173	1192	1198	1206	1213
83	755	845	957	1016	1053	1092	1125	1154	1175	1194	1200	1208	1215
84	758	849	960	1018	1056	1094	1127	1156	1177	1196	1202	1209	1216

a. Each entry is the lowest Scaled Score for that grade and percentile.

Table 60: Unified Scaled Score to Percentile Rank Conversions[a]

PR	Grade Placement												
	K	1	2	3	4	5	6	7	8	9	10	11	12
85	761	853	963	1021	1058	1096	1129	1159	1180	1198	1204	1211	1217
86	764	857	966	1023	1060	1099	1131	1161	1182	1201	1206	1212	1220
87	767	861	969	1026	1063	1101	1134	1164	1185	1204	1209	1214	1221
88	770	864	972	1029	1065	1104	1136	1166	1188	1206	1211	1216	1223
89	774	868	975	1032	1068	1108	1140	1169	1191	1209	1214	1217	1225
90	779	873	978	1035	1071	1110	1143	1172	1194	1213	1216	1220	1227
91	784	879	982	1038	1074	1113	1146	1174	1196	1215	1219	1223	1230
92	789	885	986	1041	1078	1116	1149	1177	1199	1216	1221	1225	1233
93	794	891	991	1044	1081	1120	1153	1180	1202	1219	1224	1228	1237
94	799	899	996	1048	1086	1123	1156	1183	1204	1223	1228	1232	1242
95	805	907	1002	1053	1090	1127	1160	1187	1208	1226	1232	1237	1246
96	813	917	1008	1058	1097	1132	1164	1190	1212	1230	1236	1241	1251
97	826	928	1015	1066	1104	1139	1170	1196	1218	1237	1243	1249	1260
98	844	943	1026	1074	1114	1149	1177	1203	1224	1241	1250	1255	1268
99	872	969	1041	1090	1127	1162	1188	1215	1237	1258	1267	1273	1284

a. Each entry is the lowest Scaled Score for that grade and percentile.

Table 61: Percentile Rank to Normal Curve Equivalent Conversions

PR	NCE	PR	NCE	PR	NCE	PR	NCE
1	1.0	26	36.5	51	50.5	76	64.9
2	6.7	27	37.1	52	51.1	77	65.6
3	10.4	28	37.7	53	51.6	78	66.3
4	13.1	29	38.3	54	52.1	79	67.0
5	15.4	30	39.0	55	52.6	80	67.7
6	17.3	31	39.6	56	53.2	81	68.5
7	18.9	32	40.1	57	53.7	82	69.3
8	20.4	33	40.7	58	54.2	83	70.1
9	21.8	34	41.3	59	54.8	84	70.9
10	23.0	35	41.9	60	55.3	85	71.8
11	24.2	36	42.5	61	55.9	86	72.8
12	25.3	37	43.0	62	56.4	87	73.7
13	26.3	38	43.6	63	57.0	88	74.7
14	27.2	39	44.1	64	57.5	89	75.8
15	28.2	40	44.7	65	58.1	90	77.0
16	29.1	41	45.2	66	58.7	91	78.2
17	29.9	42	45.8	67	59.3	92	79.6
18	30.7	43	46.3	68	59.9	93	81.1
19	31.5	44	46.8	69	60.4	94	82.7
20	32.3	45	47.4	70	61.0	95	84.6
21	33.0	46	47.9	71	61.7	96	86.9
22	33.7	47	48.4	72	62.3	97	89.6
23	34.4	48	48.9	73	62.9	98	93.3
24	35.1	49	49.5	74	63.5	99	99.0
25	35.8	50	50.0	75	64.2		

Table 62: Normal Curve Equivalent to Percentile Rank Conversion

NCE Range			NCE Range			NCE Range			NCE Range		
Low	High	PR	Low	High	PR	Low	High	PR	Low	High	PR
1.0	4.0	1	36.1	36.7	26	50.3	50.7	51	64.6	65.1	76
4.1	8.5	2	36.8	37.3	27	50.8	51.2	52	65.2	65.8	77
8.6	11.7	3	37.4	38.0	28	51.3	51.8	53	65.9	66.5	78
11.8	14.1	4	38.1	38.6	29	51.9	52.3	54	66.6	67.3	79
14.2	16.2	5	38.7	39.2	30	52.4	52.8	55	67.4	68.0	80
16.3	18.0	6	39.3	39.8	31	52.9	53.4	56	68.1	68.6	81
18.1	19.6	7	39.9	40.4	32	53.5	53.9	57	68.7	69.6	82
19.7	21.0	8	40.5	40.9	33	54.0	54.4	58	69.7	70.4	83
21.1	22.3	9	41.0	41.5	34	54.5	55.0	59	70.5	71.3	84
22.4	23.5	10	41.6	42.1	35	55.1	55.5	60	71.4	72.2	85
23.6	24.6	11	42.2	42.7	36	55.6	56.1	61	72.3	73.1	86
24.7	25.7	12	42.8	43.2	37	56.2	56.6	62	73.2	74.1	87
25.8	26.7	13	43.3	43.8	38	56.7	57.2	63	74.2	75.2	88
26.8	27.6	14	43.9	44.3	39	57.3	57.8	64	75.3	76.3	89
27.7	28.5	15	44.4	44.9	40	57.9	58.3	65	76.4	77.5	90
28.6	29.4	16	45.0	45.4	41	58.4	58.9	66	77.6	78.8	91
29.5	30.2	17	45.5	45.9	42	59.0	59.5	67	78.9	80.2	92
30.3	31.0	18	46.0	46.5	43	59.6	60.1	68	80.3	81.7	93
31.1	31.8	19	46.6	47.0	44	60.2	60.7	69	81.8	83.5	94
31.9	32.6	20	47.1	47.5	45	60.8	61.3	70	83.6	85.5	95
32.7	33.3	21	47.6	48.1	46	61.4	61.9	71	85.6	88.0	96
33.4	34.0	22	48.2	48.6	47	62.0	62.5	72	88.1	91.0	97
34.1	34.7	23	48.7	49.1	48	62.6	63.1	73	91.1	95.4	98
34.8	35.4	24	49.2	49.7	49	63.2	63.8	74	95.5	99.0	99
35.5	36.0	25	49.8	50.2	50	63.9	64.5	75			

Table 63: Grade Equivalent to ZPD Conversions

GE	ZPD Range		GE	ZPD Range		GE	ZPD Range	
	Low	High		Low	High		Low	High
0.0	0.0	1.0	4.4	3.2	4.9	8.8	4.6	8.8
0.1	0.1	1.1	4.5	3.2	5.0	8.9	4.6	8.9
0.2	0.2	1.2	4.6	3.2	5.1	9.0	4.6	9.0
0.3	0.3	1.3	4.7	3.3	5.2	9.1	4.6	9.1
0.4	0.4	1.4	4.8	3.3	5.2	9.2	4.6	9.2
0.5	0.5	1.5	4.9	3.4	5.3	9.3	4.6	9.3
0.6	0.6	1.6	5.0	3.4	5.4	9.4	4.6	9.4
0.7	0.7	1.7	5.1	3.5	5.5	9.5	4.7	9.5
0.8	0.8	1.8	5.2	3.5	5.5	9.6	4.7	9.6
0.9	0.9	1.9	5.3	3.6	5.6	9.7	4.7	9.7
1.0	1.0	2.0	5.4	3.6	5.6	9.8	4.7	9.8
1.1	1.1	2.1	5.5	3.7	5.7	9.9	4.7	9.9
1.2	1.2	2.2	5.6	3.8	5.8	10.0	4.7	10.0
1.3	1.3	2.3	5.7	3.8	5.9	10.1	4.7	10.1
1.4	1.4	2.4	5.8	3.9	5.9	10.2	4.7	10.2
1.5	1.5	2.5	5.9	3.9	6.0	10.3	4.7	10.3
1.6	1.6	2.6	6.0	4.0	6.1	10.4	4.7	10.4
1.7	1.7	2.7	6.1	4.0	6.2	10.5	4.8	10.5
1.8	1.8	2.8	6.2	4.1	6.3	10.6	4.8	10.6
1.9	1.9	2.9	6.3	4.1	6.3	10.7	4.8	10.7
2.0	2.0	3.0	6.4	4.2	6.4	10.8	4.8	10.8
2.1	2.1	3.1	6.5	4.2	6.5	10.9	4.8	10.9
2.2	2.1	3.1	6.6	4.2	6.6	11.0	4.8	11.0
2.3	2.2	3.2	6.7	4.2	6.7	11.1	4.8	11.1
2.4	2.2	3.2	6.8	4.3	6.8	11.2	4.8	11.2
2.5	2.3	3.3	6.9	4.3	6.9	11.3	4.8	11.3
2.6	2.4	3.4	7.0	4.3	7.0	11.4	4.8	11.4
2.7	2.4	3.4	7.1	4.3	7.1	11.5	4.9	11.5
2.8	2.5	3.5	7.2	4.3	7.2	11.6	4.9	11.6
2.9	2.5	3.5	7.3	4.4	7.3	11.7	4.9	11.7
3.0	2.6	3.6	7.4	4.4	7.4	11.8	4.9	11.8
3.1	2.6	3.7	7.5	4.4	7.5	11.9	4.9	11.9
3.2	2.7	3.8	7.6	4.4	7.6	12.0	4.9	12.0

Table 63: Grade Equivalent to ZPD Conversions

GE	ZPD Range		GE	ZPD Range		GE	ZPD Range	
	Low	High		Low	High		Low	High
3.3	2.7	3.8	7.7	4.4	7.7	12.1	4.9	12.1
3.4	2.8	3.9	7.8	4.5	7.8	12.2	4.9	12.2
3.5	2.8	4.0	7.9	4.5	7.9	12.3	4.9	12.3
3.6	2.8	4.1	8.0	4.5	8.0	12.4	4.9	12.4
3.7	2.9	4.2	8.1	4.5	8.1	12.5	5.0	12.5
3.8	2.9	4.3	8.2	4.5	8.2	12.6	5.0	12.6
3.9	3.0	4.4	8.3	4.5	8.3	12.7	5.0	12.7
4.0	3.0	4.5	8.4	4.5	8.4	12.8	5.0	12.8
4.1	3.0	4.6	8.5	4.6	8.5	12.9	5.0	12.9
4.2	3.1	4.7	8.6	4.6	8.6	13.0	5.0	13.0
4.3	3.1	4.8	8.7	4.6	8.7			

Table 64: Scaled Score to Instructional Reading Level Conversions[a]

IRL	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
Pre-Primer (PP): < 0.0	600	839	0	125
Primer (P): 0.0–0.9	839	856	125	160
1.0	856	861	160	169
1.1	861	865	169	177
1.2	865	870	177	186
1.3	870	874	186	195
1.4	874	879	195	204
1.5	879	883	204	213
1.6	883	888	213	221
1.7	888	893	221	230
1.8	893	897	230	239
1.9	897	902	239	248
2.0	902	906	248	257
2.1	906	911	257	267
2.2	911	916	267	276
2.3	916	921	276	285
2.4	921	925	285	294
2.5	925	930	294	305
2.6	930	935	305	316
2.7	935	940	316	326
2.8	940	945	326	337
2.9	945	949	337	347
3.0	949	955	347	360
3.1	955	960	360	370
3.2	960	965	370	380
3.3	965	971	380	395
3.4	971	976	395	408
3.5	976	981	408	424
3.6	981	987	424	440
3.7	987	992	440	452
3.8	992	998	452	463
3.9	998	1003	463	475

a. The figures in this table only apply to individual students, not groups.

Table 64: Scaled Score to Instructional Reading Level Conversions[a]

IRL	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
4.0	1003	1007	475	488
4.1	1007	1012	488	499
4.2	1012	1016	499	513
4.3	1016	1020	513	524
4.4	1020	1025	524	538
4.5	1025	1029	538	554
4.6	1029	1033	554	564
4.7	1033	1038	564	578
4.8	1038	1042	578	591
4.9	1042	1046	591	608
5.0	1046	1048	608	615
5.1	1048	1051	615	623
5.2	1051	1053	623	630
5.3	1053	1055	630	638
5.4	1055	1057	638	646
5.5	1057	1059	646	655
5.6	1059	1061	655	664
5.7	1061	1064	664	673
5.8	1064	1066	673	681
5.9	1066	1068	681	690
6.0	1068	1070	690	701
6.1	1070	1072	701	711
6.2	1072	1074	711	721
6.3	1074	1076	721	730
6.4	1076	1079	730	745
6.5	1079	1081	745	764
6.6	1081	1083	764	777
6.7	1083	1085	777	788
6.8	1085	1087	788	798
6.9	1087	1089	798	810
7.0	1089	1091	810	824

a. The figures in this table only apply to individual students, not groups.

Table 64: Scaled Score to Instructional Reading Level Conversions[a]

IRL	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
7.1	1091	1094	824	837
7.2	1094	1096	837	848
7.3	1096	1098	848	857
7.4	1098	1100	857	869
7.5	1100	1102	869	881
7.6	1102	1104	881	891
7.7	1104	1107	891	899
7.8	1107	1109	899	905
7.9	1109	1111	905	911
8.0	1111	1113	911	919
8.1	1113	1115	919	929
8.2	1115	1117	929	941
8.3	1117	1119	941	952
8.4	1119	1122	952	964
8.5	1122	1124	964	972
8.6	1124	1126	972	983
8.7	1126	1128	983	999
8.8	1128	1130	999	1022
8.9	1130	1132	1022	1042
9.0	1132	1135	1042	1059
9.1	1135	1137	1059	1076
9.2	1137	1139	1076	1098
9.3	1139	1141	1098	1111
9.4	1141	1143	1111	1130
9.5	1143	1145	1130	1147
9.6	1145	1147	1147	1161
9.7	1147	1150	1161	1172
9.8	1150	1152	1172	1182
9.9	1152	1154	1182	1193
10.0	1154	1156	1193	1206
10.1	1156	1158	1206	1216

a. The figures in this table only apply to individual students, not groups.

Table 64: Scaled Score to Instructional Reading Level Conversions[a]

IRL	Unified Scaled Score		Enterprise Scaled Score	
	Low	High	Low	High
10.2	1158	1160	1216	1223
10.3	1160	1162	1223	1232
10.4	1162	1165	1232	1244
10.5	1165	1167	1244	1254
10.6	1167	1169	1254	1263
10.7	1169	1171	1263	1275
10.8	1171	1173	1275	1289
10.9	1173	1175	1289	1295
11.0	1175	1177	1295	1301
11.1	1177	1180	1301	1307
11.2	1180	1182	1307	1312
11.3	1182	1184	1312	1316
11.4	1184	1186	1316	1318
11.5	1186	1188	1318	1321
11.6	1188	1190	1321	1323
11.7	1190	1193	1323	1326
11.8	1193	1195	1326	1328
11.9	1195	1197	1328	1329
12.0	1197	1199	1329	1331
12.1	1199	1202	1331	1333
12.2	1202	1204	1333	1336
12.3	1204	1207	1336	1338
12.4	1207	1209	1338	1341
12.5	1209	1212	1341	1342
12.6	1212	1214	1342	1343
12.7	1214	1217	1343	1344
12.8	1217	1219	1344	1345
12.9	1219	1222	1345	1345
Post-High School (PHS)	1222	1400	1345	1400

a. The figures in this table only apply to individual students, not groups.

Table 65: Relating Star Early Literacy Enterprise Scores to Star Reading Enterprise Scores

Star Early Literacy			Star Reading				
Star Early Literacy Scaled Score Range		Literacy Classification	Star Reading Enterprise Scaled Score Range		GE	ZPD Range	Recommended Assessment(s)
Low	High		Low	High			
300	505	Emergent Reader	NA	NA	NA	NA	Star Early Literacy
506	517		0	55	0.0	0.0–1.0	
518	531		56	57	0.1	0.1–1.1	
532	546		58	60	0.2	0.2–1.2	
547	562		61	62	0.3	0.3–1.3	
563	577		63	64	0.4	0.4–1.4	
578	594		65	67	0.5	0.5–1.5	
595	609		68	69	0.6	0.6–1.6	
610	626		70	72	0.7	0.7–1.7	
627	643		73	74	0.8	0.8–1.8	
644	659		75	77	0.9	0.9–1.9	
660	675	Transitional Reader SS = 675	78	80	1.0	1.0–2.0	
676	691		81	84	1.1	1.1–2.1	
692	706		85	87	1.2	1.2–2.2	
707	720		88	91	1.3	1.3–2.3	
721	733		92	98	1.4	1.4–2.4	
734	746		99	106	1.5	1.5–2.5	Star Early Literacy and Star Reading
747	759		107	125	1.6	1.6–2.6	
760	769		126	142	1.7	1.7–2.7	
770	779	Probable Reader SS = 775	143	158	1.8	1.8–2.8	
780	789		159	175	1.9	1.9–2.9	Star Reading
790	798		176	190	2.0	2.0–3.0	
799	806		191	206	2.1	2.1–3.1	
807	813		207	219	2.2	2.1–3.1	
814	820		220	235	2.3	2.2–3.2	
821	826		236	248	2.4	2.2–3.2	
827	831		249	262	2.5	2.3–3.3	
832	837		263	276	2.6	2.4–3.4	
838	841		277	289	2.7	2.4–3.4	

Table 65: Relating Star Early Literacy Enterprise Scores to Star Reading Enterprise Scores

Star Early Literacy			Star Reading				
Star Early Literacy Scaled Score Range		Literacy Classification	Star Reading Enterprise Scaled Score Range		GE	ZPD Range	Recommended Assessment(s)
Low	High		Low	High			
842	846	Probable Reader (continued)	290	305	2.8	2.5–3.5	Star Reading (continued)
847	849		306	318	2.9	2.5–3.5	
850	853		319	333	3.0	2.6–3.6	
854	856		334	346	3.1	2.6–3.7	
857	859		347	361	3.2	2.7–3.8	
860	862		362	371	3.3	2.7–3.8	
863	864		372	382	3.4	2.8–3.9	
865	866		383	398	3.5	2.8–4.0	
867	868		399	411	3.6	2.8–4.1	
869	870		412	425	3.7	2.9–4.2	
871	872		426	440	3.8	2.9–4.3	
873	873		441	452	3.9	3.0–4.4	

Table 66: Relating Star Early Literacy Unified Scale Scores to Star Reading GE Scores and ZPD Ranges

Star Early Literacy			Star Reading		Recommended Assessment(s)
SEL Literacy Classification	Unified Scaled Score Range		GE	ZPD Range	
	Low	High			
Emergent Reader	200	693	NA	NA	Star Early Literacy
	694	700	0.0	0.0–1.0	
	701	708	0.1	0.1–1.1	
	709	716	0.2	0.2–1.2	
	717	725	0.3	0.3–1.3	
	726	733	0.4	0.4–1.4	
	734	742	0.5	0.5–1.5	
	743	750	0.6	0.6–1.6	
	751	759	0.7	0.7–1.7	
	760	768	0.8	0.8–1.8	
	769	777	0.9	0.9–1.9	
Transitional Reader SS = 786	778	786	1.0	1.0–2.0	
	787	795	1.1	1.1–2.1	
	796	804	1.2	1.2–2.2	
	805	813	1.3	1.3–2.3	
	814	821	1.4	1.4–2.4	
	822	830	1.5	1.5–2.5	Star Early Literacy and Star Reading
	831	839	1.6	1.6–2.6	
	840	847	1.7	1.7–2.7	
Probable Reader SS = 852	848	855	1.8	1.8–2.8	
	856	864	1.9	1.9–2.9	Star Reading
	865	872	2.0	2.0–3.0	
	873	880	2.1	2.1–3.1	
	881	887	2.2	2.1–3.1	
	888	895	2.3	2.2–3.2	
	896	902	2.4	2.2–3.2	
	903	909	2.5	2.3–3.3	
	910	916	2.6	2.4–3.4	
	917	923	2.7	2.4–3.4	

Table 66: Relating Star Early Literacy Unified Scale Scores to Star Reading GE Scores and ZPD Ranges

Star Early Literacy			Star Reading		Recommended Assessment(s)
SEL Literacy Classification	Unified Scaled Score Range		GE	ZPD Range	
	Low	High			
Probable Reader (continued)	924	930	2.8	2.5–3.5	Star Reading (continued)
	931	936	2.9	2.5–3.5	
	937	943	3.0	2.6–3.6	
	944	949	3.1	2.6–3.7	
	950	955	3.2	2.7–3.8	
	956	961	3.3	2.7–3.8	
	962	966	3.4	2.8–3.9	
	967	972	3.5	2.8–4.0	
	973	977	3.6	2.8–4.1	
	978	982	3.7	2.9–4.2	
	983	987	3.8	2.9–4.3	
	988	992	3.9	3.0–4.4	

Appendix A: Estimated Oral Reading Fluency

Table 67: Estimated Oral Reading Fluency (Est. ORF) Given in Words Correct per Minute (WCPM) by Grade for Selected Star Reading Scale Score Units (SR SS)

SR SS	Grade			
	1	2	3	4
50	0	4	0	8
100	29	30	32	31
150	41	40	43	41
200	55	52	52	47
250	68	64	60	57
300	82	78	71	69
350	92	92	80	80
400	111	106	97	93
450	142	118	108	104
500	142	132	120	115
550	142	152	133	127
600	142	175	147	137
650	142	175	157	145
700	142	175	167	154
750	142	175	170	168
800	142	175	170	184
850–1400	142	175	170	190

Appendix B: Detailed Evidence of Star Reading Validity

The Validity chapter of this technical manual places its emphasis on summaries of Star Reading validity evidence, and on recent evidence which comes primarily from the 34-item, standards-based version of the assessment, which was introduced in 2011. However, the abundance of earlier evidence, and evidence related to the 25-item Star Reading versions, is all part of the accumulation of technical support for the validity and usefulness of Star Reading. Much of that cumulative evidence is presented in this appendix, to ensure that the historical continuity of research in support of Star Reading validity is not lost. The material that follows touches on the following list of topics:

- Relationship of Star Reading Scores to Scores on Other Tests of Reading Achievement
- Relationship of Star Reading Scores to Scores on State Tests of Accountability in Reading
- Relationship of Star Reading Enterprise Scores to Scores on Previous Versions
- Data from Post-Publication Studies
- Linking Star and State Assessments: Comparing Student- and School-Level Data
- Classification Accuracy and Screening Data Reported to The National Center on Response to Intervention (NCRTI)

Relationship of Star Reading Scores to Scores on Other Tests of Reading Achievement

During the Star Reading 2.0 norming study, schools submitted data on how their students performed on several standardized tests of reading achievement as well as their students' Star Reading results. This data included test results for more than 12,000 students from such tests as the California Achievement Test (CAT), the Comprehensive Test of Basic Skills (CTBS), the Iowa Test of Basic Skills (ITBS), the Metropolitan Achievement Test (MAT), the Stanford Achievement Test (SAT9), and several statewide tests.

Computing the correlation coefficients was a two-step process. First, where necessary, data were placed onto a common scale. If Scaled Scores were available, they could be correlated with Star Reading 2.0 Scaled Scores. However, since Percentile Ranks (PRs) are not on an equal-interval scale, when PRs were reported for the other tests, they were converted into Normal Curve Equivalents (NCEs). Scaled Scores or NCE scores were then used to compute the Pearson product-moment correlation coefficients.

In an ongoing effort to gather evidence for the validity of Star Reading scores, continual research on score validity has been undertaken. In addition to original validity data gathered at the time of initial development, numerous other studies have investigated the correlations between Star Reading tests and other external measures. In addition to gathering concurrent validity estimates, predictive validity estimates have also been investigated.

Table 68, Table 69, Table 70, and Table 71 present the correlation coefficients between the scores on the Star Reading test and each of the other tests for which data were received. Table 68 and Table 69 display “concurrent validity” data; that is, correlations between Star Reading test scores and other tests administered within a two-month time period. The date of administration ranged from spring 1999–spring 2013. More recently, data have become available for analyses regarding the predictive validity of Star Reading. Predictive validity provides an estimate of the extent to which scores on the Star Reading test predicted scores on criterion measures given at a later point in time, operationally defined as more than 2 months between the Star test (predictor) and the criterion test. Predictive validity provides an estimate of the linear relationship between Star scores and scores on tests covering a similar academic domain. Predictive correlations are attenuated by time due to the fact that students are gaining skills in the interim between testing occasions, and also by differences between the tests’ content specifications. Table 70 and Table 71 present predictive validity coefficients.

The tables are presented in two parts. Table 68 and Table 70 display validity coefficients for grades 1–6, and Table 69 and Table 71 display the validity coefficients for grades 7–12. The bottom of each table presents a grade-by-grade summary, including the total number of students for whom test data were available, the number of validity coefficients for that grade, and the average value of the validity coefficients.

The within-grade average concurrent validity coefficients for grades 1–6 varied from 0.72–0.80, with an overall average of 0.74. The within-grade average concurrent validity for grades 7–12 ranged from 0.65–0.76, with an overall average of 0.72. Predictive validity coefficients ranged from 0.69–0.72 in grades 1–6, with an average of 0.71. In grades 7–12 the predictive validity coefficients ranged from 0.72–0.87 with an average of 0.80. The other validity coefficient within-grade averages (for Star Reading 2.0 with external tests administered prior to spring 1999, Table 72 and Table 73) varied from 0.60–0.77; the overall average was 0.72.

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

Since correlation coefficients are available for many different test editions, forms, and dates of administration, many of the tests have several validity coefficients associated with them. Data were omitted from the tabulations if (a) test data quality could not be verified or (b) when sample size was very small. In general, these correlation coefficients reflect very well on the validity of the Star Reading test as a tool for placement in Reading. In fact, the correlations are similar in magnitude to the validity coefficients of these measures with each other. These validity results, combined with the supporting evidence of reliability and minimization of SEM estimates for the Star Reading test, provide a quantitative demonstration of how well this innovative instrument in reading achievement assessment performs.

Table 68: Concurrent Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Arkansas Augmented Benchmark Examination (AABE)														
AABE	S 08	SS	–	–	–	–	2,858	0.78*	2,588	0.73*	1,897	0.73*	1,176	0.75*
AIMSweb														
R-CBM	S 12	correct	15	0.65*	72	0.28*	41	0.17	44	0.48*	–	–	–	–
California Achievement Test (CAT)														
CAT	S 99	SS	93	0.80*	36	0.67*	–	–	34	0.72*	146	0.76*	–	–
CAT/5	F 10–11	SS	68	0.79*	315	0.72*	410	0.69*	302	0.71*	258	0.71*	196	0.69*
Canadian Achievement Test (CAT)														
CAT/2	F 10–11		–	–	–	–	21	0.80*	31	0.84*	23	0.75*	–	–
Colorado Student Assessment Program (CSAP)														
CSAP	S 06	SS	–	–	–	–	82	0.75*	79	0.83*	93	0.68*	280	0.80*

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 68: Concurrent Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Comprehensive Test of Basic Skills (CTBS)														
CTBS/4	S 99	NCE	–	–	–	–	–	–	18	0.81*	–	–	–	–
CTBS/A-19/20	S 99	SS	–	–	–	–	–	–	–	–	–	–	8	0.91*
Delaware Student Testing Program (DSTP) – Reading														
DSTP	S 05	SS	–	–	–	–	104	0.57*	–	–	–	–	–	–
DSTP	S 06	SS	–	–	158	0.68*	126	0.43*	141	0.62*	157	0.59*	75	0.66*
Dynamic Indicators of Basic Early Literacy Skills (DIBELS) – Oral Reading Fluency														
DIBELS	F 05	WCPM	–	–	59	0.78*	–	–	–	–	–	–	–	–
DIBELS	W 06	WCPM	61	0.87*	55	0.75*	–	–	–	–	–	–	–	–
DIBELS	S 06	WCPM	67	0.87*	63	0.71*	–	–	–	–	–	–	–	–
DIBELS	F 06	WCPM	–	–	515	0.78*	354	0.81*	202	0.72*	–	–	–	–
DIBELS	W 07	WCPM	208	0.75*	415	0.73*	175	0.69*	115	0.71*	–	–	–	–
DIBELS	S 07	WCPM	437	0.81*	528	0.70*	363	0.66*	208	0.54*	–	–	–	–
DIBELS	F 07	WCPM	–	–	626	0.79*	828	0.73*	503	0.73*	46	0.73*	–	–
Florida Comprehensive Assessment Test (FCAT)														
FCAT	S 06	SS	–	–	–	–	–	–	41	0.65*	–	–	–	–
FCAT	S 06–08	SS	–	–	–	–	10,169	0.76*	8,003	0.73*	5,474	0.73*	1,188	0.67*
Florida Comprehensive Assessment Test (FCAT 2.0)														
FCAT 2.0	S 13	SS	–	–	–	–	3,641	0.83*	3,025	0.84*	2,439	0.83*	145	0.81*
Gates–MacGinitie Reading Test (GMRT)														
GMRT/2nd Ed	S 99	NCE	–	–	21	0.89*	–	–	–	–	–	–	–	–
GMRT/L-3rd	S 99	NCE	–	–	127	0.80*	–	–	–	–	–	–	–	–
Idaho Standards Achievement Test (ISAT)														
ISAT	S 07–09	SS	–	–	–	–	3,724	0.75*	2,956	0.74*	2,485	0.74*	1,309	0.75*
Illinois Standards Achievement Test – Reading														
ISAT	S 05	SS	–	–	106	0.71*	594	0.76*	–	–	449	0.70*	–	–
ISAT	S 06	SS	–	–	–	–	140	0.80*	144	0.80*	146	0.72	–	–

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 68: Concurrent Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Iowa Test of Basic Skills (ITBS)														
ITBS–Form K	S 99	NCE	40	0.75*	36	0.84*	26	0.82*	28	0.89*	79	0.74*	–	–
ITBS–Form L	S 99	NCE	–	–	–	–	18	0.70*	29	0.83*	41	0.78*	38	0.82*
ITBS–Form M	S 99	NCE	–	–	–	–	158	0.81*	–	–	125	0.84*	–	–
ITBS–Form K	S 99	SS	–	–	58	0.74*	–	–	54	0.79*	–	–	–	–
ITBS–Form L	S 99	SS	–	–	–	–	45	0.73*	–	–	–	–	50	0.82*
Kansas State Assessment Program (KSAP)														
KSAP	S 06–08	SS	–	–	–	–	4,834	0.61*	4,045	0.61*	3,332	0.63*	1,888	0.65*
Kentucky Core Content Test (KCCT)														
KCCT	S 08–10	SS	–	–	–	–	10,776	0.60*	8,885	0.56*	7,147	0.53*	5,003	0.57*
Metropolitan Achievement Test (MAT)														
MAT–7th Ed.	S 99	NCE	–	–	–	–	–	–	46	0.79*	–	–	–	–
MAT–6th Ed.	S 99	Raw	–	–	–	–	8	0.58*	–	–	8	0.85*	–	–
MAT–7th Ed.	S 99	SS	–	–	–	–	25	0.73*	17	0.76*	21	0.76	23	0.58*
Michigan Educational Assessment Program (MEAP) – English Language Arts														
MEAP	F 04	SS	–	–	–	–	–	–	155	0.81*	–	–	–	–
MEAP	F 05	SS	–	–	–	–	218	0.76*	196	0.80*	202	0.80*	207	0.69*
MEAP	F 06	SS	–	–	–	–	116	0.79*	132	0.69*	154	0.81*	129	0.66*
Michigan Educational Assessment Program (MEAP) – Reading														
MEAP	F 04	SS	–	–	–	–	–	–	155	0.80*	–	–	–	–
MEAP	F 05	SS	–	–	–	–	218	0.77*	196	0.78*	202	0.81*	207	0.68*
MEAP	F 06	SS	–	–	–	–	116	0.75*	132	0.70*	154	0.82*	129	0.70*
Mississippi Curriculum Test (MCT2)														
MCT2	S 02	SS	–	–	–	–	–	–	155	0.80*	–	–	–	–
MCT2	S 03	SS	–	–	–	–	218	0.77*	196	0.78*	202	0.81*	207	0.68*
MCT2	S 08	SS	–	–	–	–	3,821	0.74*	3,472	0.73*	2,915	0.71*	2,367	0.68*
Missouri Mastery Achievement Test (MMAT)														
MMAT	S 99	NCE	–	–	–	–	–	–	–	–	26	0.62*	–	–

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 68: Concurrent Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 1–6[a]

			1		2		3		4		5		6	
Test Form	**Date**	**Score**	**n**	**r**	**n**	**r**	**n**	**r**	**n**	**r**	**n**	**r**	**n**	**r**
New Jersey Assessment of Skills and Knowledge (NJ ASK)														
NJ ASK	S 13	SS	–	–	–	–	1,636	0.79*	1,739	0.80*	1,486	0.82*	440	0.77*
New York State Assessment Program														
NYSTP	S 13	SS	–	–	–	–	185	0.78*	–	–	–	–	–	–
North Carolina End-of-Grade (NCEOG): Test														
	S 99	SS	–	–	–	–	–	–	–	–	85	0.79*	–	–
NCEOG	S 06–08	SS	–	–	–	–	2,707	0.80*	2,234	0.77*	1,752	0.77*	702	0.77*
Ohio Achievement Assessment (OAA)														
OAA	S 13	SS	–	–	–	–	1,718	0.72*	1,595	0.71*	1,609	0.77*	1,599	0.76*
Oklahoma Core Curriculum Test (OCCT)														
OCCT	S 06	SS	–	–	–	–	78	0.62*	92	0.58*	46	0.52*	80	0.60*
OCCT	S 13	SS	–	–	–	–	153	0.79*	66	0.79*	72	0.80*	64	0.72*
South Dakota State Test of Educational Progress (DSTEP)														
DSTEP	S 08–10	SS	–	–	–	–	2,072	0.78*	1,751	0.77*	1,409	0.80*	906	0.78*
Stanford Achievement Test (SAT)														
SAT 9th Ed.	S 99	NCE	68	0.79*	–	–	26	0.44*	–	–	–	–	86	0.65*
SAT 9th Ed.	S 99	SS	11	0.89*	18	0.89*	67	0.79*	66	0.79*	72	0.80*	64	0.72*
State of Texas Assessments of Academic Readiness Standards Test (STAAR)														
STAAR	S 12–13	SS	–	–	–	–	8,567	0.79*	7,902	0.78*	7,272	0.76*	5,697	0.78*
Tennessee Comprehensive Assessment Program (TCAP)														
TCAP	S 11	SS	–	–	–	–	62	0.66*	56	0.59*	–	–	–	–
TCAP	S 12	SS	–	–	–	–	91	0.79*	118	0.21*	81	0.64*	–	–
TCAP	S 13	SS	–	–	–	–	494	0.73*	441	0.66*	426	0.77*	–	–
TerraNova														
TerraNova	S 99	SS	–	–	61	0.72*	117	0.78*	–	–	–	–	–	–
Texas Assessment of Academic Skills (TAAS)														
TAAS	S 99	NCE	–	–	–	–	–	–	–	–	–	–	229	0.66*
Transitional Colorado Assessment Program (TCAP)														
TCAP	S 12–13	SS	–	–	–	–	3,144	0.78*	3,200	0.82*	3,186	0.81*	3,106	0.83*

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 68: Concurrent Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
West Virginia Educational Standards Test 2 (WESTEST 2)														
WESTEST 2	S 12	SS	–	–	–	–	2,949	0.76*	7,537	0.77*	5,666	0.76*	2,390	0.75*
Woodcock Reading Mastery (WRM)														
	S 99		–	–	–	–	–	–	–	–	7	0.68*	7	0.66*
Wisconsin Knowledge and Concepts Examination (WKCE)														
WKCE	F 06–10	SS					8,649	0.78*	7,537	0.77*	5,666	0.76*	2,390	0.75*

Summary

Grade(s)	All	1	2	3	4	5	6
Number of students	255,538	1,068	3,629	76,942	66,400	54,173	31,686
Number of coefficients	195	10	18	47	47	41	32
Average validity		0.80	0.73	0.72	0.72	0.74	0.72
Overall average	0.74						

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 69: Concurrent Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Arkansas Augmented Benchmark Examination (AABE)														
AABE	S 08	SS	318	0.79*	278	0.76*	–	–	–	–	–	–	–	–
California Achievement Test (CAT)														
CAT/5	S 99	NCE	–	–	–	–	59	0.65*	–	–	–	–	–	–
CAT/5	S 99	SS	124	0.74*	131	0.76*	–	–	–	–	–	–	–	–
CAT/5	F 10–11	SS	146	0.75*	139	0.79*	92	0.64*	81	0.82*	48	0.79*	39	0.73*
Colorado Student Assessment Program (CSAP)														
CSAP	S 06	SS	299	0.84*	185	0.83*	–	–	–	–	–	–	–	–
Delaware Students Testing Program (DSTP) – Reading														
DSTP	S 05	SS	–	–	–	–	–	–	112	0.78*	–	–	–	–
DSTP	S 06	SS	150	0.72*	–	–	–	–	–	–	–	–	–	–

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 69: Concurrent Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Florida Comprehensive Assessment Test (FCAT)														
FCAT	S 06	SS	–	–	74	0.65*	–	–	–	–	–	–	–	–
FCAT	S 06–08	SS	1,119	0.74*	618	0.76*	–	–	–	–	–	–	–	–
Florida Comprehensive Assessment Test (FCAT 2.0)														
FCAT 2.0	S 13	SS	158	0.83*	111	0.81*	–	–	–	–	–	–	–	–
Idaho Standards Achievement Test (ISAT)														
ISAT	S 06–08	SS	851	0.78*	895	0.71*	–	–	–	–	–	–	–	–
Illinois Standards Achievement Test (ISAT) – Reading														
ISAT	S 05	SS	–	–	157	0.73*	–	–	–	–	–	–	–	–
ISAT	S 06	SS	140	0.70*	–	–	–	–	–	–	–	–	–	–
Iowa Test of Basic Skills (ITBS)														
ITBS–K	S 99	NCE	–	–	–	–	67	0.78*	–	–	–	–	–	–
ITBS–L	S 99	SS	47	0.56*	–	–	65	0.64*	–	–	–	–	–	–
Kansas State Assessment Program (KSAP)														
KSAP	S 06–08	SS	1,147	0.70*	876	0.71*	–	–	–	–	–	–	–	–
Kentucky Core Content Test (KCCT)														
KCCT	S 08–10	SS	2,572	0.56*	1,198	0.56*	–	–	–	–	–	–	–	–
Michigan Educational Assessment Program – English Language Arts														
MEAP	F 04	SS	154	0.68*	–	–	–	–	–	–	–	–	–	–
MEAP	F 05	SS	233	0.72*	239	0.70*	–	–	–	–	–	–	–	–
MEAP	F 06	SS	125	0.79*	152	0.74*	–	–	–	–	–	–	–	–
Michigan Educational Assessment Program – Reading														
MEAP–R	F 04	SS	154	0.68*	–	–	–	–	–	–	–	–	–	–
MEAP–R	F 05	SS	233	0.72*	239	0.70*	–	–	–	–	–	–	–	–
MEAP–R	F 06	SS	125	0.79*	152	0.74*	–	–	–	–	–	–	–	–
Mississippi Curriculum Test (MCT2)														
MCT2	S 03	SS	372	0.70*	–	–	–	–	–	–	–	–	–	–
MCT2	S 08	SS	1,424	0.69*	1,108	0.72*	–	–	–	–	–	–	–	–
Missouri Mastery Achievement Test (MMAT)														
MMAT	S 99	NCE	–	–	29	0.78*	19	0.71*	–	–	–	–	–	–

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 69: Concurrent Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Spring 1999–Spring 2013, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
North Carolina End-of-Grade (NCEOG) Test														
NCEOG	S 06–08	SS	440	0.76*	493	0.74*	–	–	–	–	–	–	–	–
New Jersey Assessment of Skills and Knowledge (NJ ASK)														
NJ ASK	S 13	SS	595	0.78*	589	0.70*	–	–	–	–	–	–	–	–
Northwest Evaluation Association Levels Test (NWEA)														
NWEA-Achieve	S 99	NCE	–	–	124	0.66*	–	–	–	–	–	–	–	–
South Dakota State Test of Educational Progress (DSTEP)														
DSTEP	S 08–10	SS	917	0.78*	780	0.77*	–	–	–	–	–	–	–	–
Stanford Achievement Test (SAT)														
SAT–9th Ed.	S 99	NCE	50	0.65*	50	0.51*	–	–	–	–	–	–	–	–
SAT–9th Ed.	S 99	SS	70	0.70*	68	0.80*	–	–	–	–	–	–	–	–
State of Texas Assessments of Academic Readiness Standards Test (STAAR)														
STAAR	S 12–13	SS	5,062	.075*	4,651	0.75*	–	–	–	–	–	–	–	–
Test Achievement and Proficiency (TAP)														
TAP	S 99	NCE	–	–	–	–	6	0.42	13	0.80*	7	0.6	–	–
Texas Assessment of Academic Skills (TAAS)														
TAAS	S 99	NCE	–	–	–	–	–	–	43	0.60*	–	–	–	–
Transitional Colorado Assessment Program (TCAP)														
TCAP	S 12–13	SS	3,165	0.83*	3,106	0.83*	1,466	0.72*	–	–	–	–	–	–
West Virginia Educational Standards Test 2 (WESTEST 2)														
WESTEST 2	S 12	SS	1,612	0.76	1,396	0.75	–	–	–	–	–	–	–	–
Wisconsin Knowledge and Concepts Examination (WKCE)														
WKCE	F 06–10	SS	1,811	0.81	1,886	0.77	–	–	506	0.79	–	–	–	–
Wide Range Achievement Test 3 (WRAT3)														
WRAT3	S 99		–	–	17	0.81*	–	–	–	–	–	–	–	–

Summary

Grade(s)	All	7	8	9	10	11	12
Number of students	48,789	25,032	21,134	1,774	755	55	39
Number of coefficients	74	30	29	7	5	2	1
Average validity	–	0.74	0.73	0.65	0.76	0.70	0.73
Overall average	0.72						

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 70: Predictive Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Fall 2005–Spring 2013, Grades 1–6[a]

Test Form	Date[b]	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
AIMSweb														
R-CBM	S 12	correct	60	0.14	156	0.38*	105	0.11	102	0.52*	–	–	–	–
Arkansas Augmented Benchmark Examination (AABE)														
AABE	F 07	SS	–	–	–	–	5,255	0.79*	5,208	0.77*	3,884	0.75*	3,312	0.75*
Colorado Student Assessment Program (CSAP)														
CSAP	F 04	–	–	–	–	–	82	0.72*	79	0.77*	93	0.70*	280	0.77*
Delaware Student Testing Program (DSTP) – Reading														
DSTP	S 05	–	–	–	–	–	189	0.58*	–	–	–	–	–	–
DSTP	W 05	–	–	–	–	–	120	0.67*	–	–	–	–	–	–
DSTP	S 05	–	–	–	–	–	161	0.52*	191	0.55*	190	0.62*	–	–
DSTP	F 05	–	–	–	253	0.64*	214	0.39*	256	0.62*	270	0.59*	242	0.71*
DSTP	W 05	–	–	–	275	0.61*	233	0.47*	276	0.59*	281	0.62*	146	0.57*
Florida Comprehensive Assessment Test (FCAT)														
FCAT	F 05	–	–	–	–	–	–	–	42	0.73*	–	–	409	0.67*
FCAT	W 07	–	–	–	–	–	–	–	–	–	–	–	417	0.76*
FCAT	F 05–07	SS	–	–	–	–	25,192	0.78*	21,650	0.75*	17,469	0.75*	9,998	0.73*
Florida Comprehensive Assessment Test (FCAT 2.0)														
FCAT 2.0	S 13	SS	–	–	–	–	6,788	0.78*	5,894	0.80*	5,374	0.80*	616	0.74*
Idaho Standards Achievement Test (ISAT)														
ISAT	F 08–10	SS	–	–	–	–	8,219	0.77*	8,274	0.77*	7,537	0.76*	5,742	0.77*
Illinois Standards Achievement Test (ISAT) – Reading														
ISAT–R	F 05	–	–	–	–	–	450	0.73*	–	–	317	0.68*	–	–
ISAT–R	W 05	–	–	–	–	–	564	0.76*	–	–	403	0.68*	–	–
ISAT–R	F 05	–	–	–	–	–	133	0.73*	140	0.74*	145	0.66*	–	–
ISAT–R	W 06	–	–	–	–	–	138	0.76*	145	0.77*	146	0.70*	–	–
Iowa Assessment														
IA	F 12	SS	–	–	–	–	1,763	0.61*	1,826	0.61*	1,926	0.59*	1,554	0.64*
IA	W 12	SS	–	–	–	–	548	0.60*	661	0.62*	493	0.64*	428	0.65*
IA	S 12	SS	–	–	–	–	1,808	0.63*	1,900	0.63*	1,842	0.65*	1,610	0.63*

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

b. Dates correspond to the term and year of the predictor scores. With some exceptions, criterion scores were obtained during the same academic year. In some cases, data representing multiple years were combined. These dates are reported as a range (e.g. Fall 05–Fall 07).

Table 70: Predictive Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Fall 2005–Spring 2013, Grades 1–6[a]

Test Form	Date[b]	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Kentucky Core Content Test (KCCT)														
KCCT	F 07–09	SS	–	–	–	–	16,521	0.62*	15,143	0.57*	12,549	0.53*	9,091	0.58*
Michigan Educational Assessment Program (MEAP) – English Language Arts														
MEAP–EL	F 04	–	–	–	–	–	193	0.60*	181	0.70*	170	0.75*	192	0.66*
MEAP–EL	W 05	–	–	–	–	–	204	0.68*	184	0.74*	193	0.75*	200	0.70*
MEAP–EL	S 05	–	–	–	–	–	192	0.73*	171	0.73*	191	0.71*	193	0.62*
MEAP–EL	F 05	–	–	–	–	–	111	0.66*	132	0.71*	119	0.77*	108	0.60*
MEAP–EL	W 06	–	–	–	–	–	114	0.77*	–	–	121	0.75*	109	0.66*
Michigan Educational Assessment Program (MEAP) – Reading														
MEAP–R	F 04	–	–	–	–	–	193	0.60*	181	0.69*	170	0.76*	192	0.66*
MEAP–R	W 05	–	–	–	–	–	204	0.69*	184	0.74*	193	0.78*	200	0.70*
MEAP–R	S 05	–	–	–	–	–	192	0.72*	171	0.72*	191	0.74*	193	0.62*
MEAP–R	F 05	–	–	–	–	–	111	0.63*	132	0.70*	119	0.78*	108	0.62*
MEAP–R	W 06	–	–	–	–	–	114	0.72*	–	–	121	0.75*	109	0.64*
Mississippi Curriculum Test (MCT2)														
MCT2	F 01	–	–	–	86	0.57*	95	0.70*	97	0.65*	78	0.76*	–	–
MCT2	F 02	–	–	–	340	0.67*	337	0.67*	282	0.69*	407	0.71*	442	0.72*
MCT2	F 07	SS	–	–	–	–	6,184	0.77*	5,515	.74*	5,409	0.74*	4,426	0.68*
North Carolina End–of–Grade (NCEOG) Test														
NCEOG	F 05–07	SS	–	–	–	–	6,976	0.81*	6,531	0.78*	6,077	0.77*	3,255	0.77*
New York State Assessment Program														
NYSTP	S 13	SS	–	–	–	–	349	0.73*	–	–	–	–	–	–
Ohio Achievement Assessment														
OAA	S 13	SS	–	–	–	–	28	0.78*	41	0.52*	29	0.79*	30	0.75*
Oklahoma Core Curriculum Test (OCCT)														
OCCT	F 04	–	–	–	–	–	–	–	–	–	44	0.63*	–	–
OCCT	W 05	–	–	–	–	–	–	–	–	–	45	0.66*	–	–
OCCT	F 05	–	–	–	–	–	89	0.59*	90	0.60*	79	0.69*	84	0.63*
OCCT	W 06	–	–	–	–	–	60	0.65*	40	0.67*	–	–	–	–

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

b. Dates correspond to the term and year of the predictor scores. With some exceptions, criterion scores were obtained during the same academic year. In some cases, data representing multiple years were combined. These dates are reported as a range (e.g. Fall 05–Fall 07).

Table 70: Predictive Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Fall 2005–Spring 2013, Grades 1–6[a]

Test Form	Date[b]	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
South Dakota State Test of Educational Progress (DSTEP)														
DSTEP	F 07–09	SS	–	–	–	–	3,909	0.79*	3,679	0.78*	3,293	0.78*	2,797	0.79*
Star Reading														
Star–R	F 05	–	16,982	0.66*	42,601	0.78*	46,237	0.81*	44,125	0.83*	34,380	0.83*	23,378	0.84*
Star–R	F 06	–	25,513	0.67*	63,835	0.78*	69,835	0.81*	65,157	0.82*	57,079	0.83*	35,103	0.83*
Star–R	F 05	–	8,098	0.65*	20,261	0.79*	20,091	0.81*	18,318	0.82*	7,621	0.82*	5,021	0.82*
Star–R	F 05	–	8,098	0.55*	20,261	0.72*	20,091	0.77*	18,318	0.80*	7,621	0.80*	5,021	0.79*
Star–R	S 06	–	8,098	0.84*	20,261	0.82*	20,091	0.83*	18,318	0.83*	7,621	0.83*	5,021	0.83*
Star–R	S 06	–	8,098	0.79*	20,261	0.80*	20,091	0.81*	18,318	0.82*	7,621	0.82*	5,021	0.81*
State of Texas Assessments of Academic Readiness Standards Test (STAAR)														
STAAR	S 12–13	SS	–	–	–	–	6,132	0.81*	5,744	0.80*	5,327	0.79*	5,143	0.79*
Tennessee Comprehensive Assessment Program (TCAP)														
TCAP	S 11	SS	–	–	–	–	695	0.68*	602	0.72*	315	0.61*	–	–
TCAP	S 12	SS	–	–	–	–	763	0.70*	831	0.33*	698	0.65*	–	–
TCAP	S 13	SS	–	–	–	–	2,509	0.67*	1,897	0.63*	1,939	0.68*	431	0.65*
West Virginia Educational Standards Test 2 (WESTEST 2)														
WESTEST 2	S 12	SS	–	–	–	–	2,828	0.80*	3,078	0.73*	3,246	0.73*	3,214	0.73*
Wisconsin Knowledge and Concepts Examination (WKCE)														
WKCE	S 05–09	SS					15,706	0.75*	15,569	0.77*	13,980	0.78*	10,641	0.78*

Summary

Grade(s)	All	1	2	3	4	5	6
Number of students	1,227,887	74,887	188,434	313,102	289,571	217,416	144,477
Number of coefficients	194	6	10	49	43	47	39
Average validity		0.69	0.72	0.70	0.71	0.72	0.71
Overall average	0.71						

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.

b. Dates correspond to the term and year of the predictor scores. With some exceptions, criterion scores were obtained during the same academic year. In some cases, data representing multiple years were combined. These dates are reported as a range (e.g. Fall 05–Fall 07).

Table 71: Predictive Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Fall 2005–Spring 2013, Grades 7–12[a]

Test Form	Date[b]	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Arkansas Augmented Benchmark Examination (AABE)														
AABE	F 07	SS	2,418	0.74*	1,591	0.75*	–	–	–	–	–	–	–	–
Colorado Student Assessment Program (CSAP)														
CSAP	F 05	–	299	0.83*	185	0.83*	–	–	–	–	–	–	–	–
Delaware Student Testing Program (DSTP) – Reading														
DSTP	S 05	–	100	0.75*	143	0.63*	–	–	48	0.66*	–	–	–	–
DSTP	F 05	–	273	0.69*	247	0.70*	152	0.73*	97	0.78*	–	–	–	–
DSTP	W 05	–	–	–	61	0.64*	230	0.64*	145	0.71*	–	–	–	–
Florida Comprehensive Assessment Test (FCAT)														
FCAT	F 05	–	381	0.61*	387	0.62*	–	–	–	–	–	–	–	–
FCAT	W 07	–	342	0.64*	361	0.72*	–	–	–	–	–	–	–	–
FCAT	F 05–07	SS	8,525	0.72*	6,216	0.72*	–	–	–	–	–	–	–	–
Florida Comprehensive Assessment Test (FCAT 2.0)														
FCAT 2.0	S 13	SS	586	0.75*	653	0.78*	–	–	–	–	–	–	–	–
Idaho Standards Achievement Test (ISAT)														
ISAT	F 05–07	SS	4,119	0.76*	3,261	0.73*	–	–	–	–	–	–	–	–
Illinois Standards Achievement Test (ISAT) – Reading														
ISAT	F 05	–	173	0.51*	158	0.66*	–	–	–	–	–	–	–	–
Iowa Assessment														
IA	F 12	SS	1,264	0.60*	905	0.63*	–	–	–	–	–	–	–	–
IA	W 12	SS	118	0.66*	72	0.67*	–	–	–	–	–	–	–	–
IA	S 12	SS	1,326	0.68*	1,250	0.66*	–	–	–	–	–	–	–	–
Kentucky Core Content Test (KCCT)														
KCCT	F 07–09	SS	4,962	0.57*	2,530	0.58*	–	–	–	–	–	–	–	–

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.
b. Dates correspond to the term and year of the predictor scores. With some exceptions, criterion scores were obtained during the same academic year. In some cases, data representing multiple years were combined. These dates are reported as a range (e.g. Fall 05–Fall 07).

Table 71: Predictive Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Fall 2005–Spring 2013, Grades 7–12[a]

Test Form	Date[b]	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Michigan Educational Assessment Program (MEAP) – English Language Arts														
MEAP	F 04	–	181	0.71*	88	0.85*	–	–	–	–	–	–	–	–
MEAP	W 05	–	214	0.73*	212	0.73*	–	–	–	–	–	–	–	–
MEAP	S 05	–	206	0.75*	223	0.69*	–	–	–	–	–	–	–	–
MEAP	F 05	–	114	0.66*	126	0.66*	–	–	–	–	–	–	–	–
MEAP	W 06	–	114	0.64*	136	0.71*	–	–	–	–	–	–	–	–
MEAP	S 06	–	–	–	30	0.80*	–	–	–	–	–	–	–	–
Michigan Educational Assessment Program (MEAP) – Reading														
MEAP–R	F 04	–	181	0.70*	88	0.84*	–	–	–	–	–	–	–	–
MEAP–R	W 05	–	214	0.72*	212	0.73*	–	–	–	–	–	–	–	–
MEAP–R	S 05	–	206	0.72*	223	0.69*	–	–	–	–	–	–	–	–
MEAP–R	F 05	–	116	0.68*	138	0.66*	–	–	–	–	–	–	–	–
MEAP–R	W 06	–	116	0.68*	138	0.70*	–	–	–	–	–	–	–	–
MEAP–R	S 06	–	–	–	30	0.81*	–	–	–	–	–	–	–	–
Mississippi Curriculum Test (MCT2)														
MCT2	F 02	–	425	0.68*	–	–	–	–	–	–	–	–	–	–
MCT2	F 07	SS	3,704	0.68*	3,491	0.73*	–	–	–	–	–	–	–	–
North Carolina End–of–Grade (NCEOG) Test														
NCEOG	F 05–07	SS	2,735	0.77*	2,817	0.77*	–	–	–	–	–	–	–	–
Ohio Achievement Assessment														
OAA	S 13	SS	53	0.82*	38	0.66*	–	–	–	–	–	–	–	–
South Dakota State Test of Educational Progress (DSTEP)														
DSTEP	F 07–09	SS	2,236	0.79*	2,073	0.78*	–	–	–	–	–	–	–	–
Star Reading														
Star–R	F 05	–	17,370	0.82*	9,862	0.82*	2,462	0.82*	15,277	0.85*	1,443	0.83*	596	0.85*
Star–R	F 06	–	22,177	0.82*	19,152	0.82*	4,087	0.84*	2,624	0.85*	2,930	0.85*	2,511	0.86*
Star–R	F 05	–	5,399	0.81*	641	0.76*	659	0.89*	645	0.88*	570	0.90*	–	–
Star–R	F 05	–	5,399	0.79*	641	0.76*	659	0.83*	645	0.83*	570	0.87*	–	–
Star–R	S 06	–	5,399	0.82*	641	0.83*	659	0.87*	645	0.88*	570	0.89*	–	–
Star–R	S 06	–	5,399	0.80*	641	0.83*	659	0.85*	645	0.85*	570	0.86*		

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.
b. Dates correspond to the term and year of the predictor scores. With some exceptions, criterion scores were obtained during the same academic year. In some cases, data representing multiple years were combined. These dates are reported as a range (e.g. Fall 05–Fall 07).

Table 71: Predictive Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Fall 2005–Spring 2013, Grades 7–12[a]

Test Form	Date[b]	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
State of Texas Assessments of Academic Readiness Standards Test (STAAR)														
STAAR	S 12–13	SS	4,716	0.77*	4,507	0.76*	–	–	–	–	–	–	–	–
Tennessee Comprehensive Assessment Program (TCAP)														
TCAP	S 13	SS	332	0.81*	233	0.74*	–	–	–	–	–	–	–	–
West Virginia Educational Standards Test 2 (WESTEST 2)														
WESTEST 2	S 12	SS	2,852	0.71*	2,636	0.74*	–	–	–	–	–	–	–	–
Wisconsin Knowledge and Concepts Examination (WKCE)														
WKCE	S 05–09	SS	6,399	0.78*	5,500	0.78*			401	0.78*				

Summary

Grade(s)	All	7	8	9	10	11	12
Number of students	224,179	111,143	72,537	9,567	21,172	6,653	3,107
Number of coefficients	106	39	41	8	10	6	2
Average validity	–	0.72	0.73	0.81	0.81	0.87	0.86
Overall average	0.80						

a. * Denotes correlation coefficients that are statistically significant at the 0.05 level.
b. Dates correspond to the term and year of the predictor scores. With some exceptions, criterion scores were obtained during the same academic year. In some cases, data representing multiple years were combined. These dates are reported as a range (e.g. Fall 05–Fall 07).

Table 72: Other External Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Prior to Spring 1999, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
American Testronics														
Level C-3	Spr 98	Scaled	–	–	20	0.71*	–	–	–	–	–	–	–	–
California Achievement Test (CAT)														
/ 4	Spr 98	Scaled	–	–	16	0.82*	–	–	54	0.65*	–	–	10	0.88*
/ 5	Spr 98	Scaled	–	–	–	–	40	0.82*	103	0.85*	–	–	–	–
/ 5	Fall 98	NCE	40	0.83*	–	–	–	–	–	–	–	–	–	–
/ 5	Fall 98	Scaled	–	–	–	–	39	0.85*	–	–	–	–	–	–
Comprehensive Test of Basic Skills (CTBS)														
A-15	Fall 97	NCE	–	–	–	–	–	–	–	–	–	–	24	0.79*
/ 4	Spr 97	Scaled	–	–	–	–	–	–	–	–	31	0.61*	–	–
/ 4	Spr 98	Scaled	–	–	–	–	–	–	6	0.49	68	0.76*	–	–
A-19/20	Spr 98	Scaled	–	–	–	–	–	–	–	–	10	0.73*	–	–
A-15	Spr 98	Scaled	–	–	–	–	–	–	–	–	–	–	93	0.81*
A-16	Fall 98	NCE	–	–	–	–	–	–	–	–	–	–	73	0.67*
Degrees of Reading Power (DRP)														
	Spr 98		–	–	–	–	8	0.71*	–	–	25	0.72*	23	0.38
Gates-MacGinitie Reading Test (GMRT)														
2nd Ed., D	Spr 98	NCE	–	–	–	–	–	–	–	–	–	–	47	0.80*
L-3rd	Spr 98	NCE	–	–	31	0.69*	27	0.62*	–	–	–	–	–	–
L-3rd	Fall 98	NCE	60	0.64*	–	–	66	0.83*	–	–	–	–	–	–
Indiana Statewide Testing for Educational Progress (ISTEP)														
	Fall 98	NCE	–	–	–	–	19	0.80*	–	–	–	–	21	0.79*
Iowa Test of Basic Skills (ITBS)														
Form K	Spr 98	NCE	–	–	–	–	88	0.74*	17	0.59*	–	–	21	0.83*
Form L	Spr 98	NCE	–	–	–	–	50	0.84*	–	–	–	–	57	0.66*
Form M	Spr 98	NCE	–	–	68	0.71*	–	–	–	–	–	–	–	–
Form K	Fall 98	NCE	–	–	67	0.66*	43	0.73*	67	0.74*	28	0.81*	–	–
Form L	Fall 98	NCE	–	–	–	–	–	–	27	0.88*	6	0.97*	37	0.60*
Form M	Fall 98	NCE	–	–	65	0.81*	–	–	53	0.72*	–	–	–	–

a. Sample sizes are in the columns labeled "n."
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 72: Other External Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Prior to Spring 1999, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
Metropolitan Achievement Test (MAT)														
7th Ed.	Spr 98	NCE	–	–	–	–	–	–	29	0.67*	22	0.68*	17	0.86*
6th Ed	Spr 98	Raw	–	–	–	–	–	–	6	0.91*	–	–	5	0.67
7th Ed.	Spr 98	Scaled	–	–	48	0.75*	–	–	–	–	30	0.79*	–	–
7th Ed.	Fall 98	NCE	–	–	–	–	–	–	–	–	–	–	49	0.75*
Metropolitan Readiness Test (MRT)														
	Spr 96	NCE	–	–	–	–	5	0.81	–	–	–	–	–	–
	Spr 98	NCE	4	0.63	–	–	–	–	–	–	–	–	–	–
Missouri Mastery Achievement Test (MMAT)														
	Spr 98	Scaled	–	–	–	–	12	0.44	–	–	14	0.75*	24	0.62*
New York State Pupil Evaluation Program (P&P)														
	Spr 98		–	–	–	–	–	–	13	0.92*	–	–	–	–
North Carolina End of Grade Test (NCEOG)														
	Spr 98	Scaled	–	–	–	–	–	–	–	–	53	0.76*	–	–
NRT Practice Achievement Test (NRT)														
Practice	Spr 98	NCE	–	–	56	0.71*	–	–	–	–	–	–	–	–
Stanford Achievement Test (Stanford)														
9th Ed.	Spr 97	Scaled	–	–	–	–	–	–	–	–	68	0.65*	–	–
7th Ed.	Spr 98	Scaled	11	0.73*	7	0.94*	8	0.65	15	0.82*	7	0.87*	8	0.87*
8th Ed.	Spr 98	Scaled	8	0.94*	8	0.64	6	0.68	11	0.76*	8	0.49	7	0.36
9th Ed.	Spr 98	Scaled	13	0.73*	93	0.73*	19	0.62*	314	0.74*	128	0.72*	62	0.67*
4th Ed. 3/V	Spr 98	Scaled	14	0.76*	–	–	–	–	–	–	–	–	–	–
9th Ed.	Fall 98	NCE	–	–	–	–	45	0.89*	–	–	35	0.68*	–	–
9th Ed.	Fall 98	Scaled	–	–	88	0.60*	25	0.79*	–	–	196	0.73*	–	–
9th Ed. 2/SA	Fall 98	Scaled	–	–	–	–	103	0.69*	–	–	–	–	–	–
Tennessee Comprehensive Assessment Program (TCAP)														
	Spr 98	Scaled	–	–	30	0.75*	–	–	–	–	–	–	–	–

a. Sample sizes are in the columns labeled “n.”
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 72: Other External Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Prior to Spring 1999, Grades 1–6[a]

Test Form	Date	Score	1		2		3		4		5		6	
			n	r	n	r	n	r	n	r	n	r	n	r
TerraNova														
	Fall 97	Scaled	–	–	–	–	–	–	–	–	56	0.70*	–	–
	Spr 98	NCE	–	–	–	–	76	0.63*	–	–	–	–	–	–
	Spr 98	Scaled	–	–	94	0.50*	55	0.79*	299	0.75*	86	0.75*	23	0.59*
	Fall 98	NCE	–	–	–	–	–	–	–	–	–	–	126	0.74*
	Fall 98	Scaled	–	–	–	–	–	–	14	0.70*	–	–	15	0.77*
Wide Range Achievement Test 3 (WRAT3)														
	Fall 98		–	–	–	–	–	–	–	–	–	–	10	0.89*
Wisconsin Reading Comprehension Test														
	Spr 98		–	–	–	–	–	–	63	0.58*	–	–	–	–

Summary

Grade(s)	All	1	2	3	4	5	6
Number of students	4,289	150	691	734	1,091	871	752
Number of coefficients	95	7	14	19	16	18	21
Average validity	–	0.75	0.72	0.73	0.74	0.73	0.71
Overall average	0.73						

a. Sample sizes are in the columns labeled "n."
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 73: Other External Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Prior to Spring 1999, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
California Achievement Test (CAT)														
/ 4	Spr 98	Scaled	–	–	11	0.75*	–	–	–	–	–	–	–	–
/ 5	Spr 98	NCE	80	0.85*	–	–	–	–	–	–	–	–	–	–
Comprehensive Test of Basic Skills (CTBS)														
/ 4	Spr 97	NCE	–	–	12	0.68*	–	–	–	–	–	–	–	–
/ 4	Spr 98	NCE	43	0.84*	–	–	–	–	–	–	–	–	–	–
/ 4	Spr 98	Scaled	107	0.44*	15	0.57*	43	0.86*	–	–	–	–	–	–
A-16	Spr 98	Scaled	24	0.82*	–	–	–	–	–	–	–	–	–	–
Explore (ACT Program for Educational Planning, 8th Grade)														
	Fall 97	NCE	–	–	–	–	67	0.72*	–	–	–	–	–	–
	Fall 98	NCE	–	–	32	0.66*	–	–	–	–	–	–	–	–
Iowa Test of Basic Skills (ITBS)														
Form K	Spr 98	NCE	–	–	–	–	35	0.84*	–	–	–	–	–	–
Form K	Fall 98	NCE	32	0.87*	43	0.61*	–	–	–	–	–	–	–	–
Form K	Fall 98	Scaled	72	0.77*	67	0.65*	77	0.78*	–	–	–	–	–	–
Form L	Fall 98	NCE	19	0.78*	13	0.73*	–	–	–	–	–	–	–	–
Metropolitan Achievement Test (MAT)														
7th Ed.	Spr 97	Scaled	114	0.70*	–	–	–	–	–	–	–	–	–	–
7th Ed.	Spr 98	NCE	46	0.84*	63	0.86*	–	–	–	–	–	–	–	–
7th Ed.	Spr 98	Scaled	88	0.70*	–	–	–	–	–	–	–	–	–	–
7th Ed.	Fall 98	NCE	50	0.55*	48	0.75*	–	–	–	–	–	–	–	–
Missouri Mastery Achievement Test (MMAT)														
	Spr 98	Scaled	24	0.62*	12	0.72*	–	–	–	–	–	–	–	–
North Carolina End of Grade Test (NCEOG)														
	Spr 97	Scaled	–	–	–	–	–	–	58	0.81*	–	–	–	–
	Spr 98	Scaled	–	–	–	–	73	0.57*	–	–	–	–	–	–
PLAN (ACT Program for Educational Planning, 10th Grade)														
	Fall 97	NCE	–	–	–	–	–	–	–	–	46	0.71*	–	–
	Fall 98	NCE	–	–	–	–	–	–	104	0.53*	–	–	–	–
Preliminary Scholastic Aptitude Test (PSAT)														
	Fall 98	Scaled	–	–	–	–	–	–	–	–	78	0.67*	–	–

a. Sample sizes are in the columns labeled "n."
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Table 73: Other External Validity Data: Star Reading 2 Correlations (r) with External Tests Administered Prior to Spring 1999, Grades 7–12[a]

Test Form	Date	Score	7		8		9		10		11		12	
			n	r	n	r	n	r	n	r	n	r	n	r
Stanford Achievement Test (Stanford)														
9th Ed.	Spr 97	Scaled	–	–	–	–	–	–	–	–	–	–	11	0.90*
7th Ed.	Spr 98	Scaled	–	–	8	0.83*	–	–	–	–	–	–	–	–
8th Ed.	Spr 98	Scaled	6	0.89*	8	0.78*	91	0.62*	–	–	93	0.72*	–	–
9th Ed.	Spr 98	Scaled	72	0.73*	78	0.71*	233	0.76*	32	0.25	64	0.76*	–	–
4th Ed. 3/V	Spr 98	Scaled	–	–	–	–	–	–	55	0.68*	–	–	–	–
9th Ed.	Fall 98	NCE	92	0.67*	–	–	–	–	–	–	–	–	–	–
9th Ed.	Fall 98	Scaled	–	–	–	–	93	0.75*	–	–	–	–	70	0.75*
Stanford Reading Test														
3rd Ed.	Fall 97	NCE	–	–	–	–	5	0.81	24	0.82*	–	–	–	–
TerraNova														
	Fall 97	NCE	103	0.69*	–	–	–	–	–	–	–	–	–	–
	Spr 98	Scaled	–	–	87	0.82*	–	–	21	0.47*	–	–	–	–
	Fall 98	NCE	35	0.69*	32	0.74*	–	–	–	–	–	–	–	–
Test of Achievement and Proficiency (TAP)														
	Spr 97	NCE	–	–	–	–	–	–	–	–	36	0.59*	–	–
	Spr 98	NCE	–	–	–	–	–	–	41	0.66*	–	–	43	0.83*
Texas Assessment of Academic Skills (TAAS)														
	Spr 97	TLI	–	–	–	–	–	–	–	–	–	–	41	0.58*
Wide Range Achievement Test 3 (WRAT3)														
	Spr 98		9	0.35	–	–	–	–	–	–	–	–	–	–
	Fall 98		–	–	–	–	16	0.80*	–	–	–	–	–	–
Wisconsin Reading Comprehension Test														
	Spr 98		–	–	–	–	–	–	63	0.58*	–	–	–	–

Summary

Grade(s)	All	7	8	9	10	11	12
Number of students	3,158	1,016	529	733	398	317	165
Number of coefficients	60	18	15	10	8	5	4
Average validity	–	0.71	0.72	0.75	0.60	0.69	0.77
Overall average	0.71						

a. Sample sizes are in the columns labeled "n."
* Denotes correlation coefficients that are statistically significant at the 0.05 level.

Relationship of Star Reading Scores to Scores on State Tests of Accountability in Reading

In the US, following the passage of the No Child Left Behind Act (NCLB) in 2001, all states moved to comprehensive tests of grade level standards for purposes of accountability. This created interest in the degree to which Star Reading test scores are related to state accountability test scores. The following section provides specific information about the validity of Star scores relative to state test scores of the NCLB era. Results of concurrent and predictive validity (defined earlier) are presented here with specific results for a variety of state tests of accountability.

Table 74 and Table 75 provide a variety of concurrent and predictive validity coefficients, respectively, for grades 3–8. Numerous state accountability tests have been used in this research.

Table 74: Concurrent Validity Data: Star Reading 2 Correlations (r) with State Accountability Tests, Grades 3–8[a]

Date	Score	3		4		5		6		7		8	
		n	r	n	r	n	r	n	r	n	r	n	r
Colorado Student Assessment Program													
Spr 06	Scaled	82	0.75*	79	0.83*	93	0.68*	280	0.80*	299	0.84*	185	0.83*
Delaware Student Testing Program—Reading													
Spr 05	Scaled	104	0.57*	–	–	–	–	–	–	–	–	–	–
Spr 06	Scaled	126	0.43*	141	0.62*	157	0.59*	75	0.66*	150	0.72	–	–
Florida Comprehensive Assessment Test													
Spr 06	SSS	–	–	41	0.65*	–	–	–	–	–	–	74	0.65*
Illinois Standards Achievement Test—Reading													
Spr 05	Scaled	594	0.76*	–	–	449	0.70*	–	–	–	–	157	0.73*
Spr 06	Scaled	140	0.80*	144	0.80*	146	0.72*	–	–	140	0.70*	–	–
Michigan Educational Assessment Program—English Language Arts													
Fall 04	Scaled	–	–	155	0.81*	–	–	–	–	154	0.68*	–	–
Fall 05	Scaled	218	0.76*	196	0.80*	202	0.80*	207	0.69*	233	0.72*	239	0.70*
Fall 06	Scaled	116	0.79*	132	0.69*	154	0.81*	129	0.66*	125	0.79*	152	0.74*

a. Sample sizes are in the columns labeled “n.”
* Denotes correlation coefficients that are statistically significant ($p < 0.05$).

Table 74: Concurrent Validity Data: Star Reading 2 Correlations (r) with State Accountability Tests, Grades 3–8[a]

Date	Score	3		4		5		6		7		8	
		n	r	n	r	n	r	n	r	n	r	n	r
Michigan Educational Assessment Program—Reading													
Fall 04	Scaled	–	–	155	0.80*	–	–	–	–	156	0.68*	–	–
Fall 05	Scaled	218	0.77*	196	0.78*	202	0.81*	207	0.68*	233	0.71*	239	0.69*
Fall 06	Scaled	116	0.75*	132	0.70*	154	0.82*	129	0.70*	125	0.86*	154	0.72*
Mississippi Curriculum Test													
Spr 02	Scaled	148	0.62*	175	0.66*	81	0.69*	–	–	–	–	–	–
Spr 03	Scaled	389	0.71*	359	0.70*	377	0.70*	364	0.72*	372	0.70*	–	–
Oklahoma Core Curriculum Test													
Spr 06	Scaled	78	0.62*	92	0.58*	46	0.52*	80	0.60*	–	–	–	–

Summary

Grades	All	3	4	5	6	7	8
Number of students	11,045	2,329	1,997	2,061	1,471	1,987	1,200
Number of coefficients	61	12	13	11	8	10	7
Average validity	–	0.72	0.73	0.73	0.71	0.74	0.73
Overall validity	0.73						

a. Sample sizes are in the columns labeled "n."
* Denotes correlation coefficients that are statistically significant ($p < 0.05$).

Table 75: Predictive Validity Data: Star Reading Scaled Scores Predicting Later Performance for Grades 3–8 on Numerous State Accountability Tests[a]

Predictor Date	Criterion Date[b]	3		4		5		6		7		8	
		n	r	n	r	n	r	n	r	n	r	n	r
Colorado Student Assessment Program													
Fall 05	Spr 06	82	0.72*	79	0.77*	93	0.70*	280	0.77*	299	0.83*	185	0.83*
Delaware Student Testing Program—Reading													
Fall 04	Spr 05	189	0.58*	–	–	–	–	–	–	–	–	–	–
Win 05	Spr 05	120	0.67*	–	–	–	–	–	–	–	–	–	–
Spr 05	Spr 06	161	0.52*	191	0.55*	190	0.62*	–	–	100	0.75*	143	0.63*
Fall 05	Spr 06	214	0.39*	256	0.62*	270	0.59*	242	0.71*	273	0.69*	247	0.70*
Win 05	Spr 06	233	0.47*	276	0.59*	281	0.62*	146	0.57*	–	–	61	0.64*
Florida Comprehensive Assessment Test													
Fall 05	Spr 06	–	–	42	0.73*	–	–	409	0.67*	381	0.61*	387	0.62*
Win 07	Spr 07	–	–	–	–	–	–	417	0.76*	342	0.64*	361	0.72*
Illinois Standards Achievement Test—Reading													
Fall 04	Spr 05	450	0.73*	–	–	317	0.68*	–	–	–	–	–	–
Win 05	Spr 05	564	0.76*	–	–	403	0.68*	–	–	–	–	–	–
Fall 05	Spr 06	133	0.73*	140	0.74*	145	0.66*	–	–	173	0.51*	158	0.66*
Win 06	Spr 06	138	0.76*	145	0.77*	146	0.70*	–	–	–	–	–	–
Michigan Educational Assessment Program—English Language Arts													
Fall 04	Fall 05[P]	193	0.60*	181	0.70*	170	0.75*	192	0.66*	181	0.71*	88	0.85*
Win 05	Fall 05[P]	204	0.68*	184	0.74*	193	0.75*	200	0.70*	214	0.73*	212	0.73*
Spr 05	Fall 05[P]	192	0.73*	171	0.73*	191	0.71*	193	0.62*	206	0.75*	223	0.69*
Fall 05	Fall 06[P]	111	0.66*	132	0.71*	119	0.77*	108	0.60*	114	0.66*	126	0.66*
Win 06	Fall 06[P]	114	0.77*	–	–	121	0.75*	109	0.66*	114	0.64*	136	0.71*
Spr 06	Fall 06[P]	–	–	–	–	–	–	–	–	–	–	30	0.80*
Michigan Educational Assessment Program—Reading													
Fall 04	Fall 05[P]	193	0.60*	181	0.69*	170	0.76*	192	0.66*	181	0.70*	88	0.84*
Win 05	Fall 05[P]	204	0.69*	184	0.74*	193	0.78*	200	0.70*	214	0.72*	212	0.73*
Spr 05	Fall 05[P]	192	0.72*	171	0.72*	191	0.74*	193	0.62*	206	0.72*	223	0.69*
Fall 05	Fall 06[P]	111	0.63*	132	0.70*	119	0.78*	108	0.62*	116	0.68*	138	0.66*
Win 06	Fall 06[P]	114	0.72*	–	–	121	0.75*	109	0.64*	116	0.68*	138	0.70*
Spr 06	Fall 06[P]	–	–	–	–	–	–	–	–	–	–	30	0.81*

a. Grade given in the column signifies the grade within which the Predictor variable was given (as some validity estimates span contiguous grades).
b. [P] indicates a criterion measure was given in a subsequent grade from the predictor.
* Denotes significant correlation ($p < 0.05$).

Table 75: Predictive Validity Data: Star Reading Scaled Scores Predicting Later Performance for Grades 3–8 on Numerous State Accountability Tests[a]

Predictor Date	Criterion Date[b]	3		4		5		6		7		8	
		n	r	n	r	n	r	n	r	n	r	n	r
Mississippi Curriculum Test													
Fall 01	Spr 02	95	0.70*	97	0.65*	78	0.76*	–	–	–	–	–	–
Fall 02	Spr 03	337	0.67*	282	0.69*	407	0.71*	442	0.72*	425	0.68*	–	–
Oklahoma Core Curriculum Test													
Fall 04	Spr 05	–	–	–	–	44	0.63*	–	–	–	–	–	–
Win 05	Spr 05	–	–	–	–	45	0.66*	–	–	–	–	–	–
Fall 05	Spr 06	89	0.59*	90	0.60*	79	0.69*	84	0.63*	–	–	–	–
Win 06	Spr 06	60	0.65*	40	0.67*	–	–	–	–	–	–	–	–

Summary

Grades	All	3	4	5	6	7	8
Number of students	22,018	4,493	2,974	4,086	3,624	3,655	3,186
Number of coefficients	119	24	19	23	17	17	19
Average validity	–	0.66	0.68	0.70	0.68	0.69	0.70
Overall validity	0.68						

a. Grade given in the column signifies the grade within which the Predictor variable was given (as some validity estimates span contiguous grades).
b. [P] indicates a criterion measure was given in a subsequent grade from the predictor.
* Denotes significant correlation ($p < 0.05$).

Relationship of Star Reading Enterprise Scores to Scores on Previous Versions

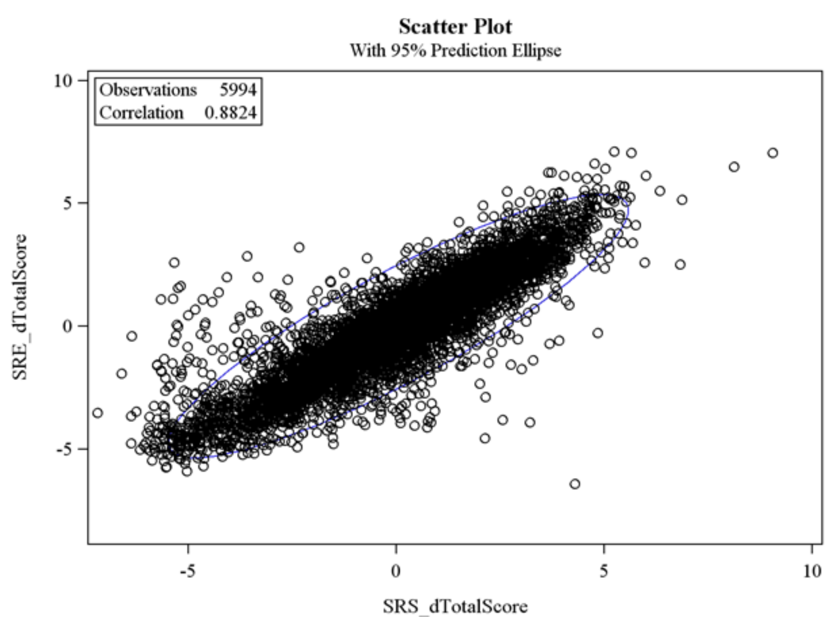
The 34-item version of Star Reading represents a significant departure from previous versions of Star. It is not a replacement for earlier versions; instead, it presents an alternative approach to reading assessment. Unlike previous Star Reading versions, which were primarily designed as measures only of reading comprehension, the 34-item version of Star Reading, simply referred to as Star Reading, is a standards-based assessment which measures a wide variety of reading skills. In addition to this substantial change in content from previous versions, Star Reading tests are also longer, and as a result have greater measurement precision and reliability.

Star Reading was released for use in June 2011. In the course of its development, Star Reading was administered to thousands of students who also took previous versions. The correlations between Star Reading and previous versions of Star Reading provide validity evidence of their own. To the extent that those correlations are high, they would provide evidence that the current Star Reading and previous versions are measuring the same or highly similar underlying attributes, even though they are dissimilar in content and measurement precision. Table 76 displays data on the correlations between Star Reading and scores on two previous versions: classic versions of Star Reading (which includes versions 2.0 through 4.3) and Star Reading Progress Monitoring (version 4.4.) Both of those Star Reading versions are 25-item versions that are highly similar to one another, differing primarily in terms of the software that delivers them; for all practical purposes, they may be considered alternate forms of Star Reading.

Table 76: Correlations of Star Reading with Scores on Star Reading Classic and Star Reading Progress Monitoring Tests

Grade	Star Reading Classic Versions		Star Reading Progress Monitoring Version	
	N	r	N	r
1	810	0.73	539	0.87
2	1,762	0.81	910	0.85
3	2,830	0.81	1,140	0.83
4	2,681	0.81	1,175	0.82
5	2,326	0.80	919	0.82
6	1,341	0.85	704	0.84
7	933	0.76	349	0.81
8	811	0.80	156	0.85
9	141	0.76	27	0.75
10	107	0.79	20	0.84
11	84	0.87	6	0.94
12	74	0.78	5	0.64
All Grades Combined	13,979	0.87	5,994	0.88

Figure 7: Scatterplot of Star Reading and Star Reading Progress Monitoring Test Scores for 5,994 Students Tested in June and July 2011



Data from Post-Publication Studies

Subsequent to publication of Star Reading 2.0 in 1999, additional external validity data became available, both from users of the assessment and from special studies conducted by Renaissance Learning and others. This section provides a summary of results of a doctoral dissertation examining the relationship of Star Reading to scores on a leading nationally standardized reading assessment, the Stanford Achievement Test (SAT9), and a major reading state test, the California Standards Test (CST).

Predictive Validity: Correlations with SAT9 and the California Standards Tests

A doctoral dissertation (Bennicoff-Nan, 2002) studied the validity of Star Reading as a predictor of student's scores in a California school district on the California Standards Test (CST) and the Stanford Achievement Tests, Ninth Edition (SAT9), the reading accountability tests mandated by the State of California. At the time of the study, those two tests were components of the California Standardized Testing and Reporting Program. The study involved analysis of test scores of more than 1,000 school children in four grades in a rural central California school district; 83 percent of students in the district were eligible for free and reduced lunch and 30 percent were identified as having limited English proficiency.

Bennicoff-Nan's dissertation addressed a number of different research questions. For purposes of this technical manual, we are primarily interested in the correlations between Star Reading 2 with SAT9 and CST scores. Those correlations are displayed by grade in Table 77.

Table 77: Correlations of Star Reading 2.0 Scores with SAT9 and California Standards Test Scores, by Grade

Grade	SAT9 Total Reading	CST English and Language Arts
3	0.82	0.78
4	0.83	0.81
5	0.83	0.79
6	0.81	0.78

In summary, the average correlation between Star Reading and SAT9 was 0.82. The average correlation with CST was 0.80. These values are evidence of the validity of Star Reading for predicting performance on both norm-referenced reading tests such as the SAT9, and criterion-referenced accountability measures such as the CST. Bennicoff-Nan concluded that Star Reading was "a time and labor effective" means of progress monitoring in the classroom, as well as suitable for program evaluation and monitoring student progress toward state accountability goals.

Linking Star and State Assessments: Comparing Student- and School-Level Data

With an increasingly large emphasis on end-of-the-year summative state tests, many educators seek out informative and efficient means of gauging student performance on state standards—especially those hoping to make instructional decisions before the year-end assessment date.

For many teachers, this is an informal process in which classroom assessments are used to monitor student performance on state standards. While this may be helpful, such assessments may be technically inadequate when compared to more standardized measures of student performance.

Recently the assessment scale associated with Star Reading has been linked to the scales used by virtually every state summative reading or ELA test in the US. Linking Star Reading assessments to state tests allows educators to reliably predict student performance on their state assessment using Star Reading scores. More specifically, it places teachers in a position to identify

- which students are on track to succeed on the year-end summative state test, and
- which students might need additional assistance to reach proficiency.

Educators using Star Reading assessments can access Star Performance Reports that allow access to students' Pathway to Proficiency. These reports indicate whether individual students or groups of students (by class, grade, or demographic characteristics) are likely to be on track to meet a particular state's criteria for reading proficiency. In other words, these reports allow instructors to evaluate student progress toward proficiency and make data-based instructional decisions well in advance of the annual state tests. Additional reports automatically generated by Star Reading help educators screen for later difficulties and progress monitor students' responsiveness to interventions.

An overview of two methodologies used for linking Star Reading to state assessments is provided in the following section.

Methodology Comparison

Recently, Renaissance Learning has developed linkages between Star Reading Scaled Scores and scores on the accountability tests of a number of states. Depending on the kind of data available for such linking, these linkages have been accomplished using one of two different methods. One method used student-level data, where both Star and state test scores were available for the same students. The other method used school-level data; this method was applied when approximately 100% of students in a school had taken Star Reading, but individual students' state test scores were not available.

Student-Level Data

Using individual data to link scores between distinct assessments is commonly used when student-level data are readily available for both assessments. In this case, the distribution of standardized scores on one test (e.g. percentile ranks) may be compared to the distribution of standardized scores on another test in an effort to establish concordance. Recently, the release of individual state test data for linking purposes allowed for the comparison of Star assessments to state test scores for several states. Star test comparison scores were obtained within an eight-week window around the median state test date (+/–4 weeks).

Typically, states classify students into one of three, four, or five performance levels on the basis of cut scores (e.g. Below Basic, Basic, Proficient, or Advanced). After each testing period, a distribution of students falling into each of these categories will always exist (e.g. 30% in Basic, 25% in Proficient, etc.). Because Star data were available for the same students who completed the state test, the distributions could be linked via equipercentile linking analysis (see Kolen & Brennan, 2004) to scores on the state test. This process creates tables of approximately equivalent scores on each assessment, allowing for the lookup of Star scale scores that correspond to the cut scores for different performance levels on the state test. For example, if 20% of students were "Below Basic" on the state test, the lowest Star cut score would be set at a score that partitioned only the lowest 20% of scores.

School-Level Data

While using student-level data is still common, obstacles associated with individual data often lead to a difficult and time-consuming process of obtaining and analyzing data. In light of the time-sensitive needs of schools, obtaining student-level data is not always an option. As an alternative, school-level data may be used in a similar manner. These data are publicly available, thus making the linking process more efficient.

School-level data were analyzed for some of the states included in the student-level linking analysis. In an effort to increase sample size, the school-level data presented here represent "projected" Scaled Scores. Each Star score was projected to the mid-point of the state test administrations window using decile-based growth norms. The growth norms are both grade- and subject-specific and are based on the growth patterns of more than one million students using Star assessments over a three-year period. Again, the linking process used for school-level data is very similar to the previously described process—the distribution of state test scores is compared to projected Star scores and using the observed distribution of state-test scores, equivalent cut scores are created for the Star assessments (the key difference being that these comparisons are made at the group level).

Accuracy Comparisons

Accuracy comparisons between student- and school-level data are particularly important given the marked resource differences between the two methods. These comparisons are presented for three states[1] in Table 78, Table 79, and Table 80. With few exceptions, results of linking using school-level data were nearly identical to student-level data on measures of specificity, sensitivity, and overall accuracy. McLaughlin and Bandeira de Mello (2002) employed similar methods in their comparison of NAEP scores and state assessment results, and this method has been used several times since then (McLaughlin & Bandeira de Mello, 2003; Bandeira de Mello, Blankenship, & McLaughlin, 2009; Bandeira et al., 2008).

In a similar comparison study using group-level data, Cronin et al. (2007) observed cut score estimates comparable to those requiring student-level data.

1. Data were available for Arkansas, Florida, Idaho, Kansas, 2Kentucky, Mississippi, North Carolina, South Dakota, and Wisconsin; however, only North Carolina, Mississippi, and Kentucky are included in the current analysis.

Table 78: Number of Students Included in Student-Level and School-Level Linking Analyses by State, Grade, and Subject

State	Grade	Reading	
		Student	School
NC	3	2,707	4,923
	4	2,234	4,694
	5	1,752	2,576
	6	702	2,604
	7	440	2,530
	8	493	1,814
MS	3	3,821	6,786
	4	3,472	7,915
	5	2,915	8,327
	6	2,367	7,861
	7	1,424	6,133
	8	1,108	4,004
KY	3	10,776	2,625
	4	8,885	4,010
	5	7,147	4,177
	6	5,003	2,848
	7	2,572	2,778
	8	1,198	1,319

Table 79: Comparison of School Level and Student Level Classification Diagnostics for Reading/ Language Arts

State	Grade	Sensitivity[a]		Specificity[b]		False + Rate[c]		False – Rate[d]		Overall Rate	
		Student	School	Student	School	Student	School	Student	School	Student	School
NC	3	89%	83%	75%	84%	25%	16%	11%	17%	83%	83%
	4	90%	81%	69%	80%	31%	20%	10%	19%	82%	81%
	5	90%	77%	69%	83%	31%	17%	10%	23%	81%	80%
	6	85%	85%	75%	75%	25%	25%	15%	15%	81%	81%
	7	84%	76%	77%	82%	23%	18%	16%	24%	80%	79%
	8	83%	79%	74%	74%	26%	26%	17%	21%	79%	76%
MS	3	66%	59%	86%	91%	14%	9%	34%	41%	77%	76%
	4	71%	68%	87%	88%	13%	12%	29%	32%	79%	79%
	5	70%	68%	84%	85%	16%	15%	30%	32%	78%	78%
	6	67%	66%	84%	84%	16%	16%	33%	34%	77%	77%
	7	63%	66%	88%	86%	12%	14%	37%	34%	79%	79%
	8	69%	72%	86%	85%	14%	15%	31%	28%	79%	80%
KY	3	91%	91%	49%	50%	51%	50%	9%	9%	83%	83%
	4	90%	86%	46%	59%	54%	41%	10%	14%	81%	80%
	5	88%	81%	50%	65%	50%	35%	12%	19%	79%	77%
	6	89%	84%	53%	63%	47%	37%	11%	16%	79%	79%
	7	86%	81%	56%	66%	44%	34%	14%	19%	77%	76%
	8	89%	84%	51%	63%	49%	37%	11%	16%	79%	78%

a. Sensitivity refers to the proportion of correct positive predictions.
b. Specificity refers to the proportion of negatives that are correctly identified (e.g. student will not meet a particular cut score).
c. False + rate refers to the proportion of students incorrectly identified as “at-risk.”
d. False – rate refers to the proportion of students incorrectly identified as *not* “at-risk.”

Table 80: Comparison of Differences Between Achieved and Forecasted Performance Levels in Reading/Language Arts (Forecast % – Achieved %)

State	Grade	Student	School	Student	School	Student	School	Student	School
NC		**Level I**		**Level II**		**Level III**		**Level IV**	
	3	–6.1%	–1.1%	2.0%	1.1%	3.6%	–0.8%	0.4%	0.9%
	4	–3.9%	–2.0%	–0.1%	1.3%	4.3%	0.4%	–0.3%	0.2%
	5	–5.1%	–1.9%	–0.7%	2.4%	8.1%	–0.7%	–2.3%	0.2%
	6	–2.1%	0.2%	0.8%	–0.4%	3.2%	–11.5%	–2.0%	11.7%
	7	–6.4%	–0.9%	2.9%	–0.4%	6.3%	–0.7%	–2.8%	2.0%
	8	–4.9%	–3.0%	3.0%	0.4%	5.1%	2.3%	–3.1%	0.3%
MS		**Minimal**		**Basic**		**Proficient**		**Advanced**	
	3	5.2%	14.1%	3.9%	0.5%	–6.1%	–13.4%	–3.0%	–1.2%
	4	5.6%	10.9%	0.2%	–3.1%	–3.0%	–5.9%	–2.8%	–1.8%
	5	4.2%	12.6%	0.4%	–6.7%	–2.7%	–7.2%	–1.9%	1.3%
	6	1.9%	6.2%	2.0%	–1.5%	–3.8%	–7.1%	0.0%	2.4%
	7	5.3%	7.0%	1.1%	–2.8%	–6.3%	–5.3%	–0.2%	1.0%
	8	6.8%	5.5%	–1.7%	–2.8%	–4.6%	–4.3%	–0.5%	1.5%
KY		**Novice**		**Apprentice**		**Proficient**		**Distinguished**	
	3	–3.5%	–1.4%	0.8%	–1.4%	6.4%	3.1%	–3.7%	–0.3%
	4	–0.5%	–0.3%	–2.5%	2.9%	6.8%	–2.1%	–3.9%	–0.5%
	5	–1.6%	1.0%	–2.3%	3.7%	9.1%	–2.9%	–5.3%	–1.8%
	6	–1.5%	1.9%	–3.6%	–1.1%	7.3%	0.0%	–2.3%	–0.8%
	7	–0.9%	0.6%	–2.5%	2.5%	6.6%	–1.7%	–3.3%	–1.4%
	8	–0.1%	1.0%	–5.1%	1.1%	8.1%	–3.0%	–2.9%	0.8%

Classification Accuracy and Screening Data: NCRTI

The National Center on Response to Intervention (NCRTI) is a federally funded project whose mission includes reviewing the technical adequacy of assessments as screening tools for use in schools adopting multi-tiered systems of support (commonly known as RTI, or response to intervention). In the July 2011 review, Star Reading earned strong ratings on NCRTI's technical criteria.

When evaluating the validity of screening tools, NCRTI considered several factors:

- classification accuracy
- validity
- disaggregated validity and classification data for diverse populations

NCRTI ratings include four qualitative labels: convincing evidence, partially convincing evidence, unconvincing evidence, or data unavailable/inadequate. Please refer to Table 81 for descriptions of these labels as provided by NCRTI, as well as the scores assigned to Star Reading in each of the categories. Further descriptive information is provided within the following tables.

Table 81: NCRTI Screening Indicator Descriptions

Indicator	Description	Star Reading Score
Classification Accuracy	Classification accuracy refers to the extent to which a screening tool is able to accurately classify students into "at risk for reading disability" and "not at risk for reading disability" categories (often evidenced by AUC values greater than 0.85).	Convincing Evidence
Validity	Validity refers to the extent to which a tool accurately measures the underlying construct that it is intended to measure (often evidenced by coefficients greater than 0.70).	Convincing Evidence
Disaggregated Validity and Classification Data for Diverse Populations	Data are disaggregated when they are calculated and reported separately for specific subgroups.	Convincing Evidence

Aggregated Classification Accuracy Data

Receiver Operating Characteristic (ROC) Curves as defined by NCRTI:

"Receiver Operating Characteristic (ROC) curves are a useful way to interpret sensitivity and specificity levels and to determine related cut scores. ROC curves are a generalization of the set of potential combinations of sensitivity and specificity possible for predictors." (Pepe, Janes, Longton, Leisenring, & Newcomb, 2004)

"ROC curve analyses not only provide information about cut scores, but also provide a natural common scale for comparing different predictors that are measured in different units, whereas the odds ratio in logistic regression analysis must be interpreted according to a unit increase in the value of the predictor, which can make comparison between predictors difficult." (Pepe, et al., 2004)

"An overall indication of the diagnostic accuracy of a ROC curve is the area under the curve (AUC). AUC values closer to 1 indicate the screening measure reliably distinguishes among students with satisfactory and unsatisfactory reading performance, whereas values at .50 indicate the predictor is no better than chance." (Zhou, Obuchowski & Obushcowski, 2002)

Brief Description of the Current Sample and Procedure

Initial Star Reading classification analyses were performed using state assessment data from Arkansas, Delaware, Illinois, Michigan, Mississippi, and Kansas. Collectively these states cover most regions of the country (Central, Southwest, Northeast, Midwest, and Southeast). Both the Classification Accuracy and Cross Validation study samples were drawn from an initial pool of 79,045 matched student records covering grades 2–11. The sample used for this analysis was 49 percent female and 28 percent male, with 44 percent not responding. Twenty-eight percent of students were White, 14 percent were Black, and 2 percent were Hispanic. Lastly, 0.4 percent were Asian or Pacific Islander and 0.2 were American Indian or Alaskan Native. Ethnicity data were not provided for 55.4 percent of the sample.

A secondary analysis using data from a single state assessment was then performed. The sample used for this analysis was 42,771 matched Star Reading and South Dakota Test of Education Progress records. The sample covered grades 3–8 and was 28 percent female and 28 percent male. Seventy-one percent of students were White and 26 percent were American Indian or Alaskan Native. Lastly, 1 percent were Black, and 1 percent were Hispanic and, 0.7 percent were Asian or Pacific Islander.

An ROC analysis was used to compare the performance data on Star Reading to performance data on state achievement tests. The Star Reading Scaled Scores used for analysis originated from assessments 3–11 months before the state achievement test was administered. Selection of cut scores was based on the graph of sensitivity and specificity versus the Scaled Score. For each grade, the Scaled Score chosen as the cut point was equal to the score where sensitivity and specificity intersected. The classification analyses, cut points and outcome measures are outlined in Table 82. When collapsed across ethnicity, AUC values were all greater than 0.80. Descriptive notes for other values represented in the table are provided in the table footnote.

Table 82: Classification Accuracy in Predicting Proficiency on State Achievement Tests in Seven States[a]

Statistic[b]	Initial Analysis		Secondary Analysis	
	Value		Value	
False Positive Rate	0.2121		0.1824	
False Negative Rate	0.2385		0.2201	
Sensitivity	0.7615		0.7799	
Specificity	0.7579		0.8176	
Positive Predictive Power	0.4423		0.5677	
Negative Predictive Power	0.9264		0.9236	
Overall Classification Rate	0.7586		0.8087	
	Grade	**AUC**	**Grade**	**AUC**

Table 82: Classification Accuracy in Predicting Proficiency on State Achievement Tests in Seven States[a]

Statistic[b]	Initial Analysis		Secondary Analysis	
	Value		Value	
AUC (ROC)	2	0.816		
	3	0.839	3	0.869
	4	0.850	4	0.882
	5	0.841	5	0.881
	6	0.833	6	0.883
	7	0.829	7	0.896
	8	0.843	8	0.879
	9	0.847		
	10	0.858		
	11	0.840		
Base	0.20		0.24	
	Grade	**Cut Score**	**Grade**	**Cut Score**
Cut Point	2	228		
	3	308	3	288
	4	399	4	397
	5	488	5	473
	6	540	6	552
	7	598	7	622
	8	628	8	727
	9	708		
	10	777		
	11	1,055		

a. Arkansas, Delaware, Illinois, Kansas, Michigan, Mississippi, and South Dakota.
b. The false positive rate is equal to the proportion of students incorrectly labeled “at-risk.” The false negative rate is equal to the proportion of students incorrectly labeled not “at-risk.” Likewise, sensitivity refers to the proportion of correct positive predictions while specificity refers to the proportion of negatives that are correctly identified (e.g., student will not meet a particular cut score).

Aggregated Validity Data

Table 83 provides aggregated validity values as well as concurrent and predictive validity evidence for Star Reading. Median validity coefficients ranged from 0.68–0.84.

Table 83: Overall Concurrent and Predictive Validity Evidence for Star Reading

Type of Validity	Grade	Test	N (Range)	Coefficient	
				Range	Median
Predictive	3–6	CST	1,000+	0.78–0.81	0.80
Predictive	2–6	SAT9	44–389	0.66–0.73	0.68
Concurrent	1–8	Suffolk Reading Scale	2,694	0.78–0.88	0.84
Construct	3, 5, 7, 10	DRP	273–424	0.76–0.86	0.82
Concurrent	1–4	DIBELS Oral Reading Fluency	12,220	0.71–0.87	0.81
Predictive	1–6	State Achievement Tests	74,877–200,929	0.68–0.82	0.79
Predictive	7–12	State Achievement Tests	3,107–64,978	0.81–0.86	0.82
Concurrent	3–8	State Achievement Tests	1,200–2,329	0.71–0.74	0.73
Predictive	3–8	State Achievement Tests	2,974–4,493	0.66–0.70	0.68

Disaggregated Validity and Classification Data

Table 84 shows the disaggregated classification accuracy data for ethnic subgroups and also the disaggregated validity data.

Table 84: Disaggregated Classification and Validity Data

Classification Accuracy in Predicting Proficiency on State Achievement Tests in 6 States (Arkansas, Delaware, Illinois, Kansas, Michigan, and Mississippi): by Race/Ethnicity					
	White, non-Hispanic (n = 17,567)	Black, non-Hispanic (n = 8,962)	Hispanic (n = 1,382)	Asian/Pacific Islander (n = 231)	American Indian/Alaska Native (n = 111)
False Positive Rate	0.3124	0.4427	0.3582	0.1710	0.1216
False Negative Rate	0.3762	0.1215	0.1224	0.2368	0.4054
Sensitivity	0.6238	0.8785	0.8776	0.7632	0.5946
Specificity	0.8676	0.5573	0.6418	0.8290	0.8784
Positive Predictive Power	0.5711	0.5031	0.6103	0.4677	0.7097
Negative Predictive Power	0.8909	0.8999	0.8913	0.9467	0.8125
Overall Classification Rate	0.8139	0.6658	0.7337	0.8182	0.7838

Table 84: Disaggregated Classification and Validity Data

Classification Accuracy in Predicting Proficiency on State Achievement Tests in 6 States (Arkansas, Delaware, Illinois, Kansas, Michigan, and Mississippi): by Race/Ethnicity										
	White, non-Hispanic (n = 17,567)		Black, non-Hispanic (n = 8,962)		Hispanic (n = 1,382)		Asian/Pacific Islander (n = 231)		American Indian/Alaska Native (n = 111)	
AUC (ROC)	**Grade**	**AUC**	**Grade**	**AUC**	**Grade**	**AUC**	**Grade**	**AUC**	**Grade**	**AUC**
	2	n/a	2	0.500	2	n/a	2	n/a	2	n/a
	3	0.863	3	0.828	3	0.868	3	0.913	3	0.697
	4	0.862	4	0.823	4	0.837	4	0.869	4	0.888
	5	0.853	5	0.832	5	0.839	5	0.855	5	0.919
	6	0.849	6	0.806	6	0.825	6	0.859	6	0.846
	7	0.816	7	0.784	7	0.866	7	0.904	7	0.900
	8	0.850	8	0.827	8	0.812	8	0.961	8	1.000
	9	1.000	9	0.848	9	n/a	9	n/a	9	n/a
	10	0.875	10	0.831	10	0.833	10	n/a	10	n/a
	11	0.750	11	1.000	11	n/a	11	n/a	11	n/a
Base Rate	0.2203		0.3379		0.3900		0.1645		0.333	
Cut Scores	**Grade**	**Cut Score**	**Grade**	**Cut Score**	**Grade**	**Cut Score**	**Grade**	**Cut Score**	**Grade**	**Cut Score**
	2	228	2	228	2	228	2	228	2	228
	3	308	3	308	3	308	3	308	3	308
	4	399	4	399	4	399	4	399	4	399
	5	488	5	488	5	488	5	488	5	488
	6	540	6	540	6	540	6	540	6	540
	7	598	7	598	7	598	7	598	7	598
	8	628	8	628	8	628	8	628	8	628
	9	708	9	708	9	708	9	708	9	708
	10	777	10	777	10	777	10	777	10	777
	11	1,055	11	1,055	11	1,055	11	1,055	11	1,055

Disaggregated Validity					
				Coefficient	
Type of Validity	**Age or Grade**	**Test or Criterion**	**n (range)**	**Range**	**Median**
Predictive (White)	2–6	SAT9	35–287	0.69–0.75	0.72
Predictive (Hispanic)	2–6	SAT9	7–76	0.55–0.74	0.675

References

Allington, R., & McGill-Franzen, A. (2003). Use students' summer-setback months to raise minority achievement. *Education Digest, 69*(3), 19–24.

Bennicoff-Nan, L. (2002). *A correlation of computer adaptive, norm referenced, and criterion referenced achievement tests in elementary reading.* Unpublished doctoral dissertation, The Boyer Graduate School of Education, Santa Ana, CA.

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4), 42–51.

Betebenner, D. W. (2010). New directions for student growth models. Retrieved from the National Center for the Improvement of Educational Assessment website: http://www.ksde.org/LinkClick.aspx?fileticket=UssiNoSZks8%3D&tabid=4421&mid=10564

Betebenner, D. W., & Iwaarden, A. V. (2011a). SGP: An R package for the calculation and visualization of student growth percentiles & percenitle growth trajectories [Computer Software manual]. (R package version 0.4-0.0 available at http://cran.r-project.org/web/packages/SGP/)

Betebenner, D. W. (2011b). A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories. The National Center for the Improvement of Educational Assessment. Retrieved from http://www.nj.gov/education/njsmart/performance/SGP_Technical_Overview.pdf

Borman, G. D. & Dowling, N. M. (2004). *Testing the Reading Renaissance program theory: A multilevel analysis of student and classroom effects on reading achievement.* University of Wisconsin-Madison.

Bracey, G. (2002). Summer loss: The phenomenon no one wants to deal with. *Phi Delta Kappan, 84*(1), 12–13.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Bulut, O., & Cormier, D. C. (2018). Validity evidence for progress monitoring with Star Reading: Slope estimates, administration frequency, and number of data points. *Frontiers in Education, 3(68)*, 1–12.

Campbell, D., & Stanley, J. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally & Company.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Boston: Houghton Mifflin Company.

Deno, S. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*(3), 184–192.

Diggle, P., Heagerty, P., Liang, K., & Zeger, S. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford: Oxford University Press.

Duncan, T., Duncan, S., Strycker, L., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Gickling, E. E., & Havertape, S. (1981). *Curriculum-based assessment (CBA).* Minneapolis, MN: School Psychology Inservice Training Network.

Gickling, E. E., & Thompson, V. E. (2001). Putting the learning needs of children first. In B. Sornson (Ed.). *Preventing early learning failure.* Alexandria, VA: ASCD.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Holmes, C. T., & Brown, C. L. (2003). *A controlled evaluation of a total school improvement process, School Renaissance.* University of Georgia. Available online: http://www.eric.ed.gov/PDFS/ED474261.pdf.

Johnson, M. S., Kress, R. A., & Pikulski, J. J. (1987). *Informal reading inventories.* Newark, DE: International Reading Association.

Kirk, R. (1995). *Experimental Design: Procedures for the behavioral sciences* (3rd ed.). New York: Brooks/Cole Publishing Company.

Kolen, M., & Brennan, R. (2004). *Test equating, scaling, and linking* (2nd ed.). New York: Springer.

McCormick, S. (1999). *Instructing students who have literacy problems* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Meyer, B., & Rice, G. E. (1984). The structure of text. In P.D. Pearson (Ed.), *Handbook on reading research* (pp. 319–352). New York: Longman.

Moskowitz, D. S., & Hershberger, S. L. (Eds.). (2002). *Modeling intraindividual variability with repeated measures data: Methods and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Multiple authors. (2002). *Modeling intraindividual variability with repeated measures data: Methods and applications.* In D. S. Moskowitz & S. L. Hershberger (eds.), Mahwah, NJ: Lawrence Erlbaum Associates.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). New York: WCB McGraw-Hill.

Pedhazur, E., & Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48(2)*.

Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. *Practical Assessment Research & Evaluation, 7(14)*.

Rudner, L. M. (2005). Expected classification accuracy. *Practical Assessment Research & Evaluation, 10(13)*.

Wyse, A. E., & Hao, S. (2012). An evaluation of item response theory classification accuracy and consistency indices. *Applied Psychological Measurement, 36(7)*, 602–624.

Index

N

O

P

R

S

T

U

V

W

Z

About Renaissance

Renaissance® transforms data about how students learn into instruments of empowerment for classroom teachers, enabling them to guide all students to achieve their full potentials. Through smart, data-driven educational technology solutions that amplify teachers' effectiveness, Renaissance helps teachers teach better, students learn better, and school administrators lead better. By supporting teachers in the classroom but not supplanting them, Renaissance solutions deliver tight learning feedback loops: between teachers and students, between assessment of skills mastery and the resources to spur progress, and between each student's current skills and future academic growth.

Renaissance

R43843.230905