RENAISSANCE®

Star Assessments™ for Spanish–Early Literacy Technical Documentation

RENAISSANCE Star Spanish®

RENAISSANCE Star Early Literacy®

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 questions): answers@renaissance.com
Email (technical questions): support@renaissance.com
Email (international support): worldsupport@renaissance.com
Website: www.renaissance.com

Copyright Notice

12/2018 SELSp

Contents

Introduction

Star Early Literacy Spanish: Screening and Progress-Monitoring Assessment

Star Early Literacy Spanish is a computer-adaptive assessment instrument designed to measure the early literacy skills of beginning Spanish readers. Star Early Literacy Spanish addresses the need to determine children's mastery of literacy concepts that are directly related to their future success as readers. Star Early Literacy Spanish assesses proficiency in two broad domains (Word Knowledge and Skills as well as Comprehension Strategies and Constructing Meaning) which include nine key early literacy sub-domains involving 34 different sets of skills or concepts. Star Early Literacy Spanish was designed explicitly to be used to assess children in kindergarten through grade 2. However, throughout its research and development, it was administered satisfactorily to children from pre-kindergarten through grade 3. In many cases, it will be suitable for teachers' use in assessing pre-kindergarten students and/or students in grade 3 and beyond.

Early childhood education programs abound in the US. Whether federally funded Head First, Even Start and Head Start programs, public preschools administered by local school districts, or private programs that are typically associated with parochial schools, the importance of assessing early literacy skills cannot be overstated. The continued ability to assess these skills during the early primary grades will enable teachers to intervene early in the formal learning process. Research supports successful early intervention as the single best predictor for future academic success, particularly in the critical areas of reading and language acquisition.

Star Early Literacy Spanish is distinguished from other assessments of early literacy in three ways. First, it is computer-administered, requiring a minimum of oversight by the teacher; its use of computer graphics, audio instructions, and computerized, automatic dictation of instructions and test questions means that most children can take the test without teacher assistance. Second, its administration is computer-adaptive, which means the content and difficulty levels of the assessment are tailored to each student's performance. Third, it is brief; each assessment administers just 27 test items and takes an average of less than ten minutes.

Unlike many assessments, Star Early Literacy Spanish is designed specifically for repeated administration throughout the school year. Star Early Literacy Spanish incorporates an automated database that records the results of each

assessment and makes them available immediately in reports of the status and growth of individual students and classes as a whole.

Star Early Literacy Spanish Purpose

Star Early Literacy Spanish is designed to provide teachers with both norm-referenced and criterion-referenced scores that will help in planning instruction and monitoring the progress of each student. Star Early Literacy Spanish supports regular assessments on a variety of literacy skills throughout the school year. This will enable teachers to easily track progress and adjust instruction based on students' current needs.

Students are expected to develop a variety of early literacy skills as they progress from pre-kindergarten through third grade. This progression reflects both the home literacy environment and educational interventions. The development of these skills is not, however, continuously upward. Students sometimes learn a skill, forget it, and relearn it, a cycle that is perfectly normal. Many well-established tests are available that test early literacy skills at a point in time, but few are designed to regularly assess a child's status at different stages through this important growth period.

Regular assessment can provide teachers with timely information concerning student understanding of literacy concepts and will prove more useful than one-time assessments. Regular assessment will also help teachers determine weekly classroom activities that will introduce students to new skills, provide them with practice so they can improve existing skills, and review skills that students may have forgotten.

Star Early Literacy Spanish is designed for regular assessment of Spanish literacy skills and concepts in kindergarten through second grade students. In many cases, its use will be appropriate in pre-kindergarten as well as in grade 3 and beyond. Star Early Literacy Spanish provides teachers with immediate feedback that will highlight instructional needs and enable teachers to target literacy instruction in order to improve the overall literacy skills of their students by some measurable means.

Star Early Literacy Spanish:

- Assesses the early literacy skills of pre-kindergarten through third grade students.
- Identifies specific areas of strength and weakness in the sub-domains and skills assessed by the program.
- Identifies students who may be at risk for later reading failure.

- Provides teachers with the following:
  - information that can be used for goal setting and outcome assessment
  - measurable information regarding individual and class literacy skills
  - timely and accurate information that can be used to plan literacy instruction and intervention
  - a tool that enables them to capture a comprehensive picture of student literacy skills in nine sub-domains
- Helps teachers monitor student progress based on the specific literacy needs of each student.

Design of Star Early Literacy Spanish

One of the fundamental design decisions concerning Star Early Literacy Spanish involved the choice of how to administer the test. The primary advantage of using computer software to administer Star Early Literacy Spanish tests is the ability to tailor each student's test based on his or her responses to previous items.

Paper-and-pencil or fixed tests are obviously far different from this: 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 in Star Early Literacy Spanish, effectively customize the test to each student's achievement level.

Adaptive Branching offers significant advantages in terms of test reliability, testing time, and student motivation. Reliability improves over paper-and-pencil or fixed-form tests because the test difficulty matches 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 fixed-form 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 design decision concerning Star Early Literacy Spanish involved the choice of the content and format of items for the test. Its content spans two domains and nine sub-domains of early literacy skills and abilities,

ranging from general readiness to vocabulary, and includes four of the five key areas of reading instruction recommended by the National Reading Panel report: phonemic awareness, phonics, vocabulary, and text comprehension. The format of the test is engaging to young children, using graphics, animation, and recorded voice to present instructions, practice, and the test items themselves.

For these reasons, Star Early Literacy Spanish's test design and item format provide a valid procedure to assess pre-reading skills and sentence and paragraph-level comprehension and to identify a student's literacy classification. Data and information presented in this documentation reinforce this.

Test Interface

The test interface for Star Early Literacy Spanish was designed to be simple, appealing to young school children, and effective. Every test question begins with dictated instructions by means of audio recordings. Additionally, every question is presented in a graphic display format. The student can replay the instructions at will; instructions will replay automatically after a measured time interval if there is no action by the student. All questions are in multiple-choice format with three response alternatives.

Students select their answers by:

- If using the keyboard, students press one of the three keys (**1**, **2**, or **3**) and then press the **Enter** key (or the **return** key on Macintosh computers).
- If using the mouse or trackpad, students select their answers by moving the cursor, then clicking the arrow button.
- If using a tablet, students tap their answer choice; then, they tap the arrow button.

Pretest Instructions

Prior to the test session itself, a brief demonstration video introduces Star Early Literacy Spanish to the student. It presents instructions in Spanish on what to expect, how to use the mouse, keyboard, or tablet, and how to answer the multiple-choice test questions.

Hands-On Exercise

To ensure that every student understands how to use the mouse, keyboard, or tablet, a short hands-on exercise precedes the assessment. The tutorial tests one of three abilities:

1. The student's ability to move the mouse or trackpad pointer to a target, to click on the target, and to click the arrow button,

2. The student's ability to press the correct key on the keyboard to choose his or her answer, and to remember to press **Enter** to move on to the next question, *or*
3. The student's ability to tap his or her answer on a tablet and to remember to tap the arrow button to move on to the next question.

Students must demonstrate proficiency in using the relevant response device before the test will proceed. A student must correctly respond to three hands-on exercise questions in a row in order to "test out" of the hands-on exercise. To correctly respond to a question, the student must have no more than one incorrect key press or off-target click (not including the Listen button) and must select the target object within five seconds after the audio instructions are through playing. When software detects that the student is having difficulty using the mouse or keyboard, the student will be instructed to ask the teacher for help.

Practice Session

After satisfactory completion of the hands-on exercise, a short practice test precedes the assessment itself. As soon as a student has answered three of five practice questions correctly, the program takes the student into the actual Star Early Literacy Spanish test. Even the youngest students should be able to answer the practice questions correctly. If the student has not successfully answered three questions in the first set of five, a second set of five practice questions is presented. Only after the student has passed the practice test does the actual test begin. Otherwise, Star Early Literacy Spanish will halt the testing session and tell the student to ask the teacher for help.

Adaptive Branching/Test Length

Star Early Literacy Spanish's branching control uses a proprietary approach somewhat more complex than the simple Rasch maximum information Item Response Theory (IRT) model. The Star Early Literacy Spanish approach was designed to yield reliable test results by adjusting item difficulty to the responses of the individual being tested while striving to minimize student frustration.

In order to minimize student frustration, the first administration of Star Early Literacy Spanish begins with items that have difficulty levels substantially below what a typical student at a given age and grade level can handle. On average, about 90 percent of students will be able to answer the first item correctly.

After the first two items, Star Early Literacy Spanish strikes a balance between student motivation and measurement efficiency by tailoring the choice of test items such that students answer an average of 75 percent of items correctly. On the second and subsequent administrations, Star Early Literacy Spanish begins testing the student at the level of his or her most recent score, again adjusting the difficulty of the early items to avoid frustration.

Once the testing session is underway, Star Early Literacy Spanish administers 27 items of varying difficulty based on the student's responses; this is sufficient information to obtain a reliable Scaled Score and to estimate the student's proficiency in all of the literacy sub-domains assessed. The average length of time to complete a Star Early Literacy Spanish test (not including the pretest instructions) is 10 minutes, with a standard deviation of approximately 3 minutes as shown in Table 1 below. Most students will be able to complete a Star Early Literacy Spanish test in under 16 minutes, including pretest instructions, and almost all will be able to do so in less than 19 minutes.

Table 1: Average and Percentiles of Total Time in Minutes to Complete Star Early Literacy Spanish Assessment During the 2017–2018 School Year

Grade	Sample Size	Mean	Standard Deviation	5th Percentile	50th Percentile	95th Percentile	99th Percentile
Pre-K	1,213	7.43	1.85	5.47	6.98	11.1	14.15
0	70,477	7.86	2.39	5.42	7.18	12.63	16.33
1	65,365	9.05	2.7	5.68	8.52	14.22	17.25
2	10,356	9.96	2.67	6.48	9.5	14.92	17.88
3	1,993	9.97	2.7	6.57	9.52	15.18	18.63

Test Repetition

Star Early Literacy Spanish 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 recommends assessing students only as frequently as necessary to get the data needed. Schools that use Star Spanish for screening purposes typically administer it three times per year. Currently, Star Early Literacy Spanish may be administered as frequently as 3 times a year for progress monitoring purposes.

The Star Early Literacy Spanish item bank contains more than 770 items, so students can test periodically without getting the same questions more than once. Star Early Literacy Spanish 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 30-day period.

Item Time Limits

The Star Early Literacy Spanish test has time limits for individual items that are based on latency data obtained during item calibration. These time limits are imposed not to ensure rapid responses, but to keep the test moving should the student become distracted and to ensure test security should the student walk away. Items that time out are counted as incorrect responses. (If the student selects the correct response, but does not press enter or return and does not select the arrow button by time-out, the item is counted as a correct response.) Students have up to 45 seconds to answer each hands-on exercise question, up to 60 seconds to answer each practice question, and up to 90 seconds to answer each actual test question for both operational and uncalibrated items. Star Early Literacy Spanish also provides the 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. When the extend time option is on, students have up to 45 seconds to answer each hands-on exercise question, up to 180 seconds to answer each practice question, and up to 270 seconds to answer each actual test question for both operational and uncalibrated items. When a student has only 15 seconds remaining for a given item (10 seconds during the hands-on exercise), a chime sounds, a clock appears, and the student is reminded to select an answer.

Repeating the Instructions

If a student wants to repeat the instructions for the current item, he or she can do so by pressing the **E** key on the keyboard or clicking/tapping (if using a tablet) the **Escuchar** button on the screen. This will cause the instructions to be replayed. The instructions will also be replayed automatically if there is no student action within a preset interval following the initial play of the instructions. The length of that interval varies according to item type, with a longer interval in the case of items that require more time for the student to process them.

Test Security

Star Early Literacy Spanish includes many features intended to provide adequate security to protect the content of the test and to maintain the confidentiality of the test results.

Split-Application Model

In the Star Early Literacy Spanish software, when students log in, they do not have access to the same functions that teachers, administrators, and other personnel can access. Students are allowed to test, but they have no other features available in Star Early Literacy Spanish; therefore they have no access to confidential information. When teachers and administrators log in, they can manage student and class information, set preferences, register students for testing, and create informative reports about student test performance.

Individualized Tests

Using Adaptive Branching, every Star Early Literacy Spanish 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 codes, it is practically impossible to read the Star Early Literacy Spanish 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 group: these groups have different names depending on which platform the user's site is on.

- For customers on the original platform, the groups are called user groups, and there are seven of them (district administrator, district staff, school administrator, school staff, teachers, students, and parents). Each user group is granted a specific set of *capabilities*.
- For customers who have been migrated to the new Renaissance Growth Platform, the groups are called user permission groups, and there are six of them (district level administrator, district dashboard owner, district staff, school level administrator, school staff, and teacher). Each user permission group is granted a specific set of *user permissions*.

Each capability or user permission corresponds to one or more tasks that can be performed in the program. The capabilities/user permissions for these groups can be changed, and they 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 Early Literacy Spanish tests from unauthorized computers (such as home computers). For more information on student access security, see https://help.renaissance.com/RP/SettingSecurityOptions or https://help2.renaissance.com/setup/22509.

The security of the Star Early Literacy Spanish 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, and staff) must enter a user name and password before they can access the data and create reports. Parents on original sites 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 Early Literacy Spanish software.

Test Monitoring/Password Entry

Monitoring of student tests is another useful security feature of Star Early Literacy Spanish. Test monitoring is implemented using the Testing Password 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 Early Literacy Spanish can do a lot to provide specific measures of test security, the real line of defense against unauthorized access or misuse of the program is the users' responsibility. Educators need to be careful not to leave the program running unattended and to monitor 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 Early Literacy Spanish's security and the quality and the validity of its scores.

Content and Item Development

Star Early Literacy Spanish consists of 770+ operational items (as of November of 2018) that align to a set of early literacy skills derived from current research of Spanish-language literacy acquisition and exemplary state standards. The test administers 27 items per session, with test items aligned to 2 broad blueprint domains which are organized into 9 blueprint sub-domains as follows:

Blueprint Domains:

- Word Knowledge and Skills
- Comprehension Strategies and Constructing Meaning

Blueprint Sub-Domains:

- Alphabetic Principle
- Concept of Word
- Visual Discrimination
- Phonemic Awareness
- Phonics
- Structural Analysis
- Vocabulary
- Sentence-Level Comprehension
- Paragraph-Level Comprehension

Blueprint skill sets and blueprint domains in Star Early Literacy Spanish are based on extensive analysis of curriculum materials and state standards and have been reviewed by Spanish-language early learning consultants.

This structure encompasses four of the five critical areas of reading instruction identified by the National Reading Panel and CCSS. The one area not covered fully by Star Early Literacy Spanish is fluency, a reading behavior that is best assessed by other means. However, fluency is well-known to be highly correlated with other reading skills, such as comprehension and using context to determine word meaning, both of which are assessed in Star Early Literacy for Spanish.

Development History

The first stage of Star Early Literacy Spanish content development was to identity the set of skills to be assessed. The test design for Star Early Literacy Spanish is similar to the English-language version of Star Early Literacy.

Differences between the two products appear at the level of blueprint skill and item design in skill sets where there are linguistic and semantic differences between English and Spanish. Early numeracy is not assessed in Star Early Literacy Spanish. Multiple resources were consulted to determine the set of skills most appropriate for assessing the reading development of K–3 US students, including those areas where Spanish-language reading development differs from English. These resources include but are not limited to:

- *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*. http://www.nagb.org/publications/frameworks/reading09.pdf
- Common Core State Standards Initiative (2010). Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects
- *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects, Spanish Language Version*, San Diego County Office of Education. https://commoncore-espanol.sdcoe.net/CCSS-en-Espanol/SLA-Literacy.
- Individual state standards from all 50 states.
- *Texas Essential Knowledge and Skills: Spanish Language Arts and Reading English as a Second Language*, Texas Education Agency. http://ritter.tea.state.tx.us/rules/tac/chapter128/index.html

The development of the skills list included iterative reviews by reading and assessment experts and psychometricians specializing in educational assessment. See Table 2 on page 13 for the Star Early Literacy Spanish Blueprint skills list.

The second stage of content development was to develop and calibrate Spanish-language versions of psychometrically validated Star Early Literacy English-language items for skills applicable to both English- and Spanish-language literacy acquisition. Star Early Literacy Spanish items were drawn from English-language items using the process of transadaptation. Transadaptation involves both the translation and adaptation of English-language items and the replacement of items unfit for translation/transadaptation with items written in Spanish. This process ensures that test items accurately assess the targeted skills while also being sensitive to semantic differences between Spanish and English, as well as appropriate at grade level. All transadaptation was performed by a professional

Spanish-language translation vendor and reviewed by Spanish-fluent editors at Renaissance. A strict development process was maintained to ensure quality item development.

The third and ongoing stage of content development is to develop items written directly in Spanish. All writers and editors have content-area expertise, relevant classroom experience, and native-level knowledge of Spanish and Spanish pedagogy, and they use those qualifications in determining grade-level appropriateness for each item developed. 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. A strict development process is maintained to ensure quality item development.

Test Blueprint Characteristics

Every Star Early Literacy Spanish assessment consists of items that tap knowledge and skills from as many as nine different literacy sub-domains. The items comprise several sets of skills for each sub-domain, with 36 different sets of skills in all.

Content balancing specifications, known as the test blueprint, ensure that a specific number of items from each blueprint sub-domain are administered in every test. A summary of the test blueprint for Star Early Literacy Spanish appears here.

Each Star Early Literacy Spanish test consists of 27 scored items, and a separately-specified number of uncalibrated items. Uncalibrated items do not affect a student's score, but allow Renaissance to measure the psychometric validity of assessment items before selecting items for use in the operational bank.

During a single test, with some exceptions, no more than 3 items are administered from the same blueprint skill set.

Blueprint Sub-Domain Prescriptions

For the first test a student takes during a school year, the number of items administered from each blueprint sub-domain is prescribed by grade (pre-K, K, 1, 2, 3).

Subsequent to that initial test, the prescriptions are governed by bands of scale scores on the previous test. Additionally, restrictions in the software program ensure that questions that require the ability to read are not administered to students below the first grade.

The Star Early Literacy Spanish Item Bank

Within the two Star Early Literacy Spanish blueprint domains, closely related skill sets are organized into blueprint sub-domains. The resulting hierarchical structure is domain, sub-domain, skill set, and specific skill. Each Star item is designed to assess a specific skill within the test blueprint. Table 2 displays the domains, sub-domains, skill sets, and skills.

Table 2: Hierarchical Structure of the Star Early Literacy Spanish Item Bank

Domain: Word Knowledge and Skills		
Blueprint Sub-Domain	**Blueprint Skill Set**	**Blueprint Skill**
Alphabetic Principle	Alphabetic Knowledge	Recognize lowercase letters
		Recognize uppercase letters
		Match lowercase letters with uppercase letters
		Match uppercase with lowercase letters
		Distinguish numbers from letters
	Alphabetic Sequence	Identify the letter that comes next
		Identify the letter that comes before
	Letter Sounds	Recognize sounds of lowercase letters
		Recognize sounds of uppercase letters
Concept of Word	Print Concepts: Word Length	Identify longest word
		Identify shortest word
	Print Concepts: Word Borders	Identify number of words (2–3)
	Print Concepts: Letters and Words	Differentiate words from letters
		Differentiate letters from words
Visual Discrimination	Letters	Differentiate lowercase letters
		Differentiate uppercase letters
		Differentiate lowercase letters in a mixed set
		Differentiate uppercase letters in a mixed set
	Identification and Word Matching	Identify words that are different
		Match words that are the same
		Identify words that are different from a prompt
Phonemic Awareness	Rhyming and Word Families	Match sounds within word families (named pictures)
		Match sounds within word families
		Identify rhyming words (named pictures)
		Identify nonrhyming words (named pictures)

Table 2: Hierarchical Structure of the Star Early Literacy Spanish Item Bank (Continued)

Domain: Word Knowledge and Skills		
Blueprint Sub-Domain	**Blueprint Skill Set**	**Blueprint Skill**
Phonemic Awareness (continued)	Blending Word Parts	Blend onsets and rimes
		Blend 2-syllable words
		Blend 3-syllable words
	Blending Phonemes	Blend phonemes in (VC) or (CVC) words
		Blend phonemes in single-syllable words
	Initial and Final Phonemes	Determine which word (picture) has an initial phoneme different from a prompt
		Determine which word (picture) has a different initial phoneme
		Match initial phoneme to a prompt (pictures)
		Recognize same final sounds (pictures)
		Determine which word (picture) has a final phoneme different from a prompt
		Determine which word (picture) has a different final phoneme
	Consonant Blends (PA)	Match consonant blend sounds (pictures)
	Phoneme Isolation/Manipulation	Substitute initial syllable (named pictures)
		Substitute initial syllable (unnamed pictures)
		Determine missing phoneme, initial or final
		Substitute initial syllable in a prompt (pictures)
		Substitute final syllable sound in a prompt (unnamed pictures)
	Phoneme Isolation/Manipulation (continued)	Substitute final syllable (named pictures)
		Substitute final syllable sound (unnamed pictures)
		Substitute vowel sounds (pictures)
Phonics	Vowel Sounds	Match vowel sounds to a prompt (words)
		Match vowel sounds to letters
		Decode CVC words
		Recognize vowel sounds (words)
		Distinguish vowel sounds (words)
		Decode grade-appropriate words
	Initial Consonant Sounds	Identify initial consonant sound (words)
		Identify letter for initial consonant sound (words and letters)

Table 2: Hierarchical Structure of the Star Early Literacy Spanish Item Bank (Continued)

Domain: Word Knowledge and Skills		
Blueprint Sub-Domain	**Blueprint Skill Set**	**Blueprint Skill**
Phonics (continued)	Final Consonant Sounds	Match words to a given final consonant sound
		Identify letter for a final consonant sound
	Consonant Blends (PH)	Recognize initial consonant blends (words)
		Distinguish consonant blends (words)
		Recognize word with a consonant blend in a contextual sentence
		Recognize associated spelling patterns of initial consonant blends
		Recognize associated spelling patterns of final consonant blends
	Consonant Digraphs	Identify a consonant digraph in a named word
		Identify a consonant digraph in an unnamed word
		Identify a contextual word containing a consonant digraph
		Identify correct spelling of consonant digraphs in words
	Other Vowel Sounds	Identify diphthong sounds in words
		Decode words with diphthongs and recognize associated spelling patterns
	Sound Symbol Correspondence: Consonants	Substitute initial consonants (words)
		Substitute final consonants (words)
		Substitute final consonant sound (named words)
	Sound Symbol Correspondence: Consonants (continued)	Substitute final consonant sound (unnamed words)
	Word Building	Identify words made by adding an initial consonant (unnamed words)
		Identify words made by adding an additional medial letter (unnamed words)
		Identify words made by adding an additional final letter (unnamed words)
		Identify words built by adding one letter to an audio prompt
	Sound Symbol Correspondence: Vowels	Substitute vowel sounds (words)

Table 2: Hierarchical Structure of the Star Early Literacy Spanish Item Bank (Continued)

Domain: Word Knowledge and Skills		
Blueprint Sub-Domain	**Blueprint Skill Set**	**Blueprint Skill**
Phonics (continued)	Word Families/Rhyming	Identify rhyming words (words)
		Identify nonrhyming words (words)
		Identify rhyming words (unnamed answers)
		Identify rhyming words (unnamed prompt and answers)
		Identify nonrhyming words (unnamed prompt and answers)
		Identify onset/rime in named words
		Identify onset/rime in unnamed words
		Identify sounds within word families (named words)
		Identify sounds within word families (unnamed words)
Structural Analysis	Words with Affixes	Use knowledge of common affixes to decode words
	Syllabification	Use knowledge of syllable patterns to decode words
		Decode multisyllable words
	Compound Words	Identify compound words (named words)
		Identify words that are not compounds (named words)
		Identify compound words (unnamed words)
		Identify words that are not compounds (unnamed words)
		Identify correctly formed compounds
Vocabulary	Word Facility	Match words to pictures
		Read high frequency words by sight
		Determine categorical relationships
		Understand position words
		Read grade-level sight words
		Understand multi-meaning words
	Synonyms	Identify synonyms of grade-appropriate words
		Identify synonym of a grade-appropriate word in a contextual sentence
		Match words with their synonyms in paragraph context (assisted)
		Match words with their synonyms in paragraph context (unassisted)

Table 2: Hierarchical Structure of the Star Early Literacy Spanish Item Bank (Continued)

Domain: Word Knowledge and Skills		
Blueprint Sub-Domain	**Blueprint Skill Set**	**Blueprint Skill**
Vocabulary (continued)	Antonyms	Identify antonyms of words
		Identify antonyms of words in context (assisted)
		Identify antonyms of words in context (unassisted)
Domain: Comprehension Strategies and Constructing Meaning		
Blueprint Sub-Domain	**Blueprint Skill Set**	**Blueprint Skill**
Sentence-Level Comprehension	Comprehension at the Sentence Level	Listen and identify word in context
		Read and identify word in context
Paragraph-Level Comprehension	Comprehension of Paragraphs	Identify the main topic of a text
		Listen to text and answer literal *who*, *what* questions
		Listen to text and answer *where, when, why* questions
		Read text and answer literal *who, what* questions
		Read text and answer *where, when, why* questions

Item Design Guidelines

Every assessment item was written to the following specifications:

Simplicity

Items directly address the skill in the most straightforward manner possible. Evaluators should have no difficulty deducing the exact nature of the skill set or skill being assessed by the item. Instructions should be explicit, clear, simple, and consistent from one item to the next.

Screen Layout

The testing screen should feel comfortable for the student and teacher. Background colors should be unobtrusive and relatively muted, and text and graphics should stand out clearly against the background. The item background must be the same for all items on the test.

Each item consists of a combination of audio instructions, an on-screen prompt in the form of a cloze stem containing text or graphics, and three answer choices containing letters, words, graphics, or numbers.

Text

For letter and word identification items, the type size should be large, but may become smaller for higher grades. The type size should be tied to items, so that it varies according to the developmental level of a student. Text size should always be consistent from one answer choice to the others so as not to make one answer choice stand out visually.

The student instructions for every test item are administered aurally by the computer, so there is no need for printed directions on-screen. For any items that utilize on-screen text, either as part of answer choices or for items that require reading, the type is be a sans-serif font of appropriate size.

Every effort is made to use common words as the target and distractor words in test items.

Graphics

Any graphical depictions should be easily recognized by students. Color should be functional, as opposed to decorative, and lines should be as smooth as possible. For complex graphics, such as those needed for listening comprehension, line drawings on a light background should be used. The size and placement of the graphics should be consistent throughout.

The art for correct answers and distractors should be consistent in order to avoid introducing an extraneous error source. Answer choices will primarily consist of graphics and text, but sound or animation occasionally will be needed. Art should be acceptable to a broad range of teachers, parents, and students, avoiding controversial or violent graphics of any kind.

Answer Options

All items have three answer choices. Only one of the choices is the correct answer. Answer choices are always arranged horizontally. For students using the keyboard as the means of choosing their answer, the answers have small text below each answer choice labeling A, B, and C, moving from left to right.

Distractors are chosen to provide the most common errors in recognition, matching, and comprehension tasks.

Words and artwork used in answer choices should be reused in no more than 10% of the items within a skill set, a sub-domain, or within the item bank as a whole. For example, a picture of a cat should only appear as an answer choice in no more than 10 out of 100 items in any given skill set.

Language and Pronunciation

All item directions will be delivered auditorily so as to support students that are not yet reading or emergent readers. Instructions will clearly state the item stem with the instruction for the student to "elige la…" Vocal talent should read questions with a neutral Spanish voice. Pacing should be slow enough to be clearly understood while still be engaging for students.

Language should be used consistently throughout the assessment. Standard protocols should be established for item administration that reflect consistent instructions. For example, if an item stem is repeated twice, the same repetition should be used for all items of the same type. One exception to this rule is those situations where the same item type is used across grades, and one of the factors that changes is the level of instruction provided to the student.

In Phonemic Awareness items, words should be segmented into phonemes, that is, divided into their individual sounds. As much as possible, the individual sounds should be preserved, and not distorted in any way.

In the recording of item instructions and answer sounds, the audio segments should minimize the tendency to add a vowel sound after a consonant sound, especially for unvoiced consonants, such as /p/, /k/, and /t/. For example, /p/ should not be pronounced "puh." Instead, it should be spoken in a loud whisper and in a clipped manner.

For voiced consonants that cannot be pronounced without a vowel sound, such as /b/ and /g/, the audio segments should keep the vowel sound as short as possible. Vowel sounds should be pronounced by simply lengthening the sound of the vowel.

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 and nouns, gender is also represented by pronoun. 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. Gender within a Spanish item is also not indicated by the usage of a gendered noun, assuming the noun is not referring to a person.
- **Ethnicity:** The goal is to create a balance among the following designations for US products: 17% White, 13% Black or African American,

60% Hispanic, 2% Middle Eastern, and 6% 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.

Sub-Domain Item Design

Content for Star Early Literacy Spanish approximately covers a range of items broad enough to test students from pre-kindergarten through grade 3 as well as remedial students in grade 4. There are also enough test items for assessing skills in nine sub-domains. The following sub-domains are considered essential in reading development:

1. **Alphabetic Principle (AP)**—Knowledge of letter names, alphabetic letter sequence, and the sounds associated with letters.
2. **Concept of Word (CW)**—Understanding of print concepts regarding written word length and word borders and the difference between words and letters.
3. **Visual Discrimination (VS)**—Differentiating both upper- and lowercase letters, identifying words that are different and matching words that are the same.
4. **Phonemic Awareness (PA)**—Understanding of rhyming words, ability to blend and segment word parts and phonemes, isolating and manipulating initial, final, and medial phonemes, and identifying the sounds in consonant blend.
5. **Phonics (PH)**—Understanding of vowels and vowel sounds, initial and final consonants, consonant blends and digraphs, consonant and vowel substitution, and identification of rhyming words and sounds in word families.
6. **Structural Analysis (SA)**—Understanding affixes and syllable patterns in decoding, and identification of compound words.
7. **Vocabulary (VO)**—Knowledge of high-frequency words, regular and irregular sight words, multi-meaning words, and words used to describe categorical relationships, position words, synonyms, and antonyms.
8. **Sentence-Level Comprehension (SC)**—Identification of words in context.

9. **Paragraph-Level Comprehension (PC)**—Identification of the main topic of text and ability to answer literal and inferential questions after listening to or reading text.

Readability Guidelines

The readability levels for each script within each item should not exceed the grade level of the item. Words used in scripts should be appropriate for the intended grade.

The content in Star Early Literacy Spanish is leveled to address pre-readers and beginning readers (generally children of ages 3 through 9).

Items in each of the sub-domains were designed to range from easy to difficult. This was achieved through the use of different combinations of audio and graphic elements, such as named pictures, unnamed pictures, named letters and sounds, unnamed letters and sounds, named words, and unnamed words, sentences, and paragraphs. The level of difficulty for each question was controlled through the use of graphical, textual, and audio support.

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 literary and information 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 present any ethnicity, gender, culture, economic class, or religion unfavorably.
- Items do not reference illegal activities, sinister or depressing subjects, religious activities or holidays based on religious activities, witchcraft, or unsafe activities.

Item and Scale Calibration

Background

The current version of Star Early Literacy Spanish was published in 2018 as measure of early literacy skills in Spanish. Over 2,100 items measuring pre-literacy skills were calibrated. The items were calibrated from response data collected using multiple fixed field test forms administered on the computer. This chapter summarizes the process of item and scale calibration using the Rasch model. It also introduces the Unified Scale used to report the Star Early Literacy Spanish scores.

Calibration of Initial Star Early Literacy Spanish Items

This section describes the process by which the Star Early Literacy Spanish items were calibrated on a common scale of difficulty. A later section will describe the process that has been used subsequently to calibrate new items; we call that process "dynamic calibration." Subject matter experts and editors reviewed the content of every item and recommended retaining some and rejecting others. After this item content review, over 770 items, measuring nine broad literacy areas and 34 literacy-related skills, remained as candidates for inclusion in the item bank.

In order to use the test items for computer-adaptive testing, every item had to be placed on a continuous scale of difficulty—the same scale used to select items adaptively and to score the adaptive tests. The procedures of IRT were chosen as the basis for scaling Star Early Literacy Spanish item difficulty, a process called "calibration."

IRT calibration is based on statistical analysis of response data—it requires hundreds of responses to every test item. To obtain these data, Renaissance conducted a major item Calibration Study in the spring of 2018.

Sample Description

A sample of 30,806 students in pre-kindergarten through grade 3 in over 1,140 schools in the United States participated in the Calibration Study. The sample was predominantly Hispanic. The calibration sample did not need to be nationally representative, but it did require a wide range of student abilities at each grade or age level.

Table 3 provides the number of students in each grade who participated in the calibration study.

Table 3: Number of students Tested by Grade, Star Early Literacy Spanish Calibration Study—Spring 2018

Grade Level	Number of Students Tested
Pre-K	392
K	13,746
1	12,761
2	3,107
3	800

The demographics of the calibration sample are summarized in Tables 4 and 5 below. Table 4 shows that the sample consisted primarily of Hispanic students. Table 5 shows an almost equal gender split in the calibration sample, as expected.

Table 4: Summary of the Calibration Sample by Ethnicity

Ethnicity[a]	Calibration Study
American Indian	6.1%
Asian	0.9%
Black	2.4%
Hispanic	80.0%
White	10.6%

a. There were 8,935 students with no ethnicity reported.

Table 5: Summary of the Calibration Sample by Gender

Gender	Calibration Study
Female	50.4%
Male	49.6%

The objective of the Calibration Study was to collect sufficient response data to allow IRT item parameters to be estimated for all 2,100+ Star Early Literacy Spanish items.

In support of this objective, provisions were made during forms design to facilitate expressing all IRT item parameters on a common scale. To that end, some of the test items were used as “anchor items”—items common to two or more forms that are used to facilitate linking all items to the common scale.

Two kinds of anchoring were used: 1) horizontal (form-to-form) anchoring, and 2) vertical (level-to-level) anchoring.

Horizontal anchoring: The purpose of horizontal anchoring is to place all items at a given level on the same scale, regardless of differences among the forms at that level. To accomplish that, several items appeared in all forms at a given level. These horizontal anchor items were chosen to be representative of the content domains and to be appropriate for the grade level.

Vertical anchoring: The purpose of vertical anchoring is to place items at adjacent levels on the same scale. To accomplish that, a number of items were administered at each of two adjacent levels: on grade and one grade above.

Table 6 breaks down the composition of test forms at each grade level in terms of number of test questions, as well as the number of calibration test forms at each grade level. Students answered a set of items at the current grade level as well as a number of questions one grade level below their grade.

Table 6: Calibration Test Forms Design by Grade Level, Star Early Literacy Spanish Calibration Study—Spring 2018

Grade Level	Items per Form	Number of Forms
Pre-K and K	31	54
1	35	46
2 and 3	35	11
Sum		111
x Counterbalancing Factor		2
Total Number of Forms		222

For reliable IRT scale linking, it is important for anchor items to be representative of the content of the tests they are used to anchor. To that end, the distribution of anchor items was approximately proportional to the distribution of items among the domains and skills summarized in the "Content and Item Development" chapter.

Because all Star Early Literacy Spanish test items include computerized graphics and audio, the calibration test forms were all computer-administered.

Item Response Function

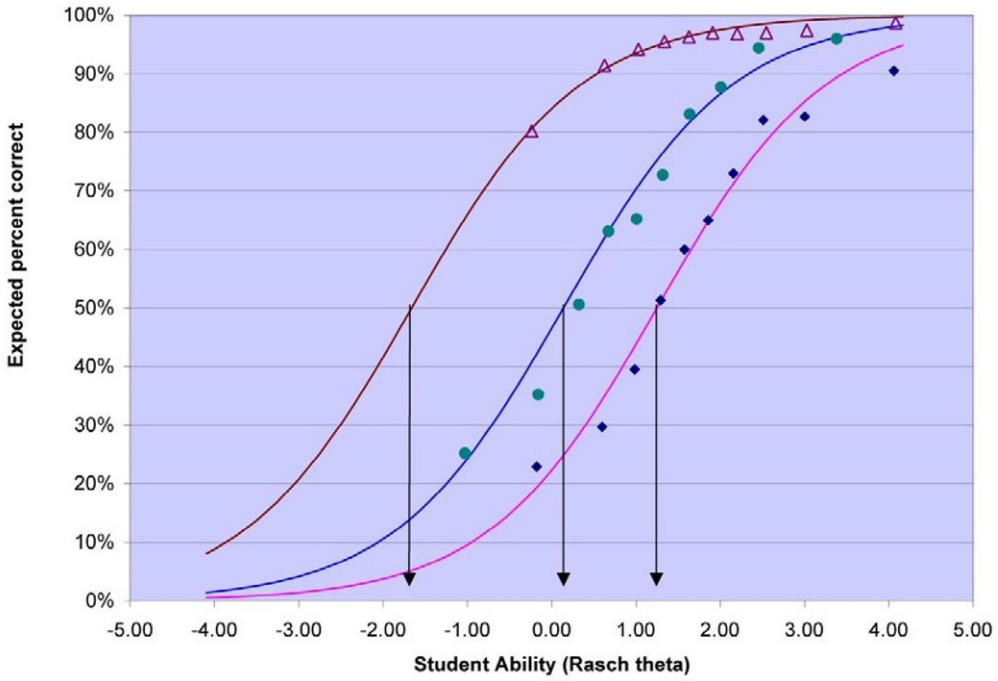
With the response data from the Calibration Study in hand, the first order of business was to calibrate the items and score the students' tests. This was done using the "Rasch model," an IRT model that expresses the probability of a correct answer as a function of the difference between the locations of the item and the student on a common scale.

Although IRT encompasses a family of mathematical models, the one-parameter (or Rasch) IRT model was selected for the Star Early Literacy Spanish data both for its simplicity and its ability to accurately model the performance of the Star Early Literacy Spanish 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 1 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.

Figure 1: Example of Item Statistics Database Presentation of Information



In Figure 1, 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 1 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.

Rasch analysis was used to determine the value of a "difficulty parameter" for every item, and to assign a score to every student. In the analysis, a number of statistical measures of item quality and model fit were calculated for each item.

Item parameter estimation and IRT scoring were accomplished using both WINSTEPS® and SAS®. Both are commercially available Rasch analysis software packages. WINSTEPS is capable of Rasch analysis of multiple test forms simultaneously. Using this capability, a concurrent calibration of all of the Star Early Literacy Spanish items was conducted. After the Winsteps calibration, more data were collected, and the final item calibration was conducted in SAS.

The principal end products of the item calibration process were the IRT item parameter estimates themselves, along with traditional indices of item difficulty (sample proportion correct) and item discriminating power (correlation coefficients between item score and the Rasch ability score).

Item Retention

Once the calibration analysis was complete, a psychometric review took place. The review evaluated both the IRT-based results and the traditional item analysis results, such as proportion correct and item-total correlations.

Reviewers evaluated each item's difficulty, discriminating power, model fit indices, statistical properties and content to identify any items that appeared unsuitable for inclusion in the Star Early Literacy Spanish adaptive testing item bank.

Of the 2,100+ items in the calibration item bank, 770+ were accepted by the psychometric review team for use in the adaptive version of Star Early Literacy Spanish. The final version of the Star Early Literacy Spanish item bank therefore contains 770+ items.

Score Scale Definition and Development

After item calibration using the Rasch IRT model, a score scale was developed for use in reporting Star Early Literacy Spanish results. Although the Rasch ability scale could be used for this purpose, a more "user-friendly" scale was preferred.[1] A system of integer numbers ranging from 200 to 1400 was chosen as the score reporting scale for Star Early Literacy Spanish.

Because there is a relationship between early literacy skills and the eventual ability to read, a decision was made to link Star Early Literacy Spanish to its counterpart Star Early Literacy English scale which is in turn linked to the Star Reading scale. This was accomplished by creating a linear link between the two Rasch scales using the common items approach as described in Kolen and Brennan (2004, pages 162–165).

Once the Star Early Literacy Spanish scores were on the same scale as the Star Early Literacy English scores, the same approach used in the Star English applications to transform scores to the Unified Scale were used to report Unified scale scores for Star Early Literacy Spanish.

More specifically, Star Early Literacy Spanish Rasch scores were linearly converted to the Star Early Literacy English Rasch scores. The Star Early Literacy English Rasch scores were then linearly transformed to the Star Reading English Rasch scale. To aid score interpretation, the final transformation equation: Unified Scale Score = INT (42.93 * Star Early Literacy Spanish Rasch Score + 958.74) was applied to report scores for Star Early Literacy Spanish.

Computer-Adaptive Test Design

In computer-adaptive tests like the Star Early Literacy Spanish 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 early literacy skills may branch to easier items in order to better estimate his or her early literacy achievement level.

High-performing students may branch to more challenging early literacy items in order to better determine the breadth of their early literacy skills and their early literacy achievement level.

1. Scores on the Rasch ability scale are expressed on the "real number" line, use decimal fractions, and can be either negative or positive. While useful for scientific and technical analysis, the Rasch ability scale does not lend itself to comfortable interpretation by teachers and lay persons.

During a Star Early Literacy Spanish test, a student may be "routed" to items at the lowest early literacy level or to items at higher early literacy levels within the overall pool of items, depending on the student's 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 Early Literacy Spanish item calibration study.

Students who have not taken a Star Early Literacy Spanish 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 Early Literacy Spanish test within the last six months, the difficulty of the first item depends on that student's previous Star Early Literacy Spanish test score information. After the administration of the initial item, and after the student has entered an answer, Star Early Literacy Spanish software estimates the student's reading ability. The software then selects the next item randomly from among all of the items available that are consistent with content constraints and most closely match the student's estimated early literacy ability.

Random selection of eligible items with difficulty values near the student's adjusted early literacy ability allows the program to avoid overexposure of test items. Items that have been administered to the same student within the past 30-day time period 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 Early Literacy Spanish computer-adaptive test.

Scoring in the Star Early Literacy Spanish Tests

Following the administration of each Star Early Literacy Spanish item, and after the student has selected an answer, an updated estimate of the student's 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 Early Literacy Spanish software uses a proprietary Maximum-Likelihood IRT estimation procedure to avoid any potential of bias in the Scaled Scores.

This approach to scoring enables Star Early Literacy Spanish 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). Unlike a conventional paper-and-pencil test, the CSEM values for the Star Early Literacy Spanish 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 Early Literacy Spanish (grades Pre-K–3). 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.

On-line Data Collection for New Item Calibration

Beginning with the 2018–2019 school year, new tests items at grade levels Pre-K through 3 are being developed and calibrated using the "dynamic calibration" method. Dynamic calibration is Renaissance's method of collecting and analyzing response data on new Star items, including Early Literacy Spanish.

Dynamic calibration allows response data on new test items to be collected during the Star testing sessions for the purpose of field testing and calibrating those items. When dynamic calibration is active, it works by embedding one or more new items at random points during a Star test. These items do not count towards the student's Star test score, but item responses are stored for later psychometric analysis. Students may take as many as three additional items per test; in some cases, no additional items will be administered. On average, this will only increase testing time by one to two minutes. The new, non-calibrated items will not count towards students' final scores, but will be analyzed in conjunction with the responses of hundreds of other students.

The response data collected on new items allows for continual evaluation of new item content and will contribute to continuous improvement in Star tests' assessment of student performance.

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 Early Literacy Spanish, 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 Early Literacy Spanish tests provide two ways to evaluate the reliability of scores: reliability coefficients, which indicate the overall precision of a set of test scores, and standard errors of measurement (SEM), which provide an index of the degree of error in test scores.

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 Early Literacy Spanish, two types of SEM are calculated: "global SEM," which is a summary of a test's measurement error, calculated for a sample or population of examinees; and "conditional SEM," CSEM. CSEM is an estimate of the measurement error in 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.

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 and the global standard error of measurement of Star Early Literacy Spanish test scores.

Generic Reliability

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

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

where σ^2_{error} is the variance of the errors of measurement, and σ^2_{total} is the variance of the test scores. In Star Early Literacy Spanish, 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.

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

where the summation is over the squared values of the reported CSEM for students *i* = 1 to *n*. In each Star Early Literacy Spanish test, the 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 a stratified random sample of the Star Early Literacy Spanish data from the 2016–2017 and 2017–2018 school years resulted in the generic reliability estimates shown in Table 7. Table 7 shows that the generic reliability estimates on the Unified scale range from 0.76 in Pre-K to 0.88 in grade 3. The overall generic reliability is 0.88. These estimates show a high degree of reliability of the Star Early Literacy Spanish scores. 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 Star Early Literacy Spanish.

Table 7: Internal Consistency, Retest Reliability, and Split Half Reliability of Star Early Literacy Spanish on the Unified Scale (Assessments Taken in the 2016–2017 and 2017–2018 School Years)

	Reliability Estimates, Unified Scale						
	Generic		Split-Half		Alternate Forms		
Grade	N	ρxx	N	ρxx	N	ρxx	Average Days Between Testing
Pre-K	1,200	0.76	1,200	0.76	300	0.49	92
K	10,000	0.80	10,000	0.78	2,500	0.55	109
1	10,000	0.83	10,000	0.81	2,500	0.73	72
2	2,500	0.85	2,500	0.85	500	0.77	68
3	500	0.88	500	0.87	100	0.75	68
Overall	24,200	0.88	24,200	0.87	5,900	0.77	88

As the data in Table 7 show, Star Early Literacy Spanish generic reliability is high, grade by grade and overall.

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 (Lord and Novick, pp. 112–113), 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 Early Literacy Spanish reliability. These split-half reliability coefficients are independent of the generic reliability approach discussed above and more firmly grounded in the item response data.

Table 7 contains split-half reliability estimates for Star Early Literacy Spanish, calculated from a stratified random sample of 2016–2017 and 2017–2018 school year data. Split-half scores based on the odd- and the even-numbered items were calculated. Adjusting the half-length assessments (13 items) to full length assessments (27 items) using the Spearman Brown formula yielded the split-half reliability estimates displayed in Table 7. As shown in the table, the split-half reliability estimates are very similar to the generic reliabilities, providing further proof of the high reliability of Star Early Literacy Spanish scores.

Alternate-Forms 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 comparable, 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 forms reliability study provided estimates of Star Early Literacy Spanish 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 Early Literacy Spanish alternate forms 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 forms reliability estimates for the Star Early Literacy Spanish test were calculated using the Star Early Literacy Spanish Unified Scaled Scores. Checks were made for valid test data on both test administrations and to remove cases of apparent motivational discrepancies.

Table 7 on page 33 includes overall and within-grade alternate forms reliability estimates, along with an indication of the average number of days between testing occasions. The average number of days between testing occasions ranged from 68–109 days. Results indicated that the overall reliability of the scores was about 0.77. The alternate forms coefficients ranged from a low of 0.49 in pre-K to a high of 0.77 in grade 2.

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 Early Literacy Spanish administration. In other words, the actual Star Early Literacy Spanish reliability is likely higher than the alternate form reliability estimates indicate.

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; smaller values of SEM or CSEM indicate greater precision.

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 30 means that if a student were tested repeatedly, his or her scores would fluctuate within 30 points of his or her first score about 68 percent of the time, and within 60 points (twice the SEM) roughly 95 percent of the time. Since reliability can also be regarded as a measure of precision, there is an inverse relationship between the reliability of a test and the standard error of measurement for the scores it produces; lower standard error of measurement results in higher reliability.

The Star Early Literacy Spanish tests differ from traditional tests in at least two respects with regard to the standard error of measurement. First, Star Early Literacy Spanish 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 Early Literacy Spanish 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 Early Literacy Spanish test is adaptive, the CSEM will tend to be lower than that of a conventional test of the same length, 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 generally similar for all students.

Table 8 contains two different sets of estimates of Star Early Literacy Spanish 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 8 are the average CSEM values observed for each grade.

Global standard error of measurement is based on the traditional SEM estimation method, using the estimated generic reliability and the variance of the test scores to estimate the SEM:

$$SEM = \sqrt{1 - \rho_{xx}}\sigma_x$$

where ρ_{XX} is the estimated generic reliability, and σ_X is the standard deviation of the observed scores (in this case, Scaled Scores).

Table 8 summarizes the distribution of CSEM values for the 2016–2017 and 2017–2018 data, overall and by grade level. The overall average CSEM was 30

scaled score units and ranged from a low of 29 in Pre-K to a high of 32 in grade 3 (Table 8).

Table 8 also shows the estimates of the global SEM. The global SEM estimates were slightly higher than the CSEM estimates. The overall SEM was 33. Across grades, the SEM ranged from a low of 29 in grade 1 to a high of 33 in grade 3.

Table 8: Star Early Literacy Spanish Standard Errors of Measurement on the Unified Scale from 2016–2017 and 2017–2018 School Year Data

Grade	Sample Size	Standard Error of Measurement, Unified Scale		
		Conditional		
		Average	Standard Deviation	Global
Pre-K	1,200	29	2.6	29
K	10,000	30	3.1	32
1	10,000	30	3.7	31
2	2,500	31	5.4	32
3	500	32	6.7	33
Overall	24,200	30	3.8	33

Validity

Test validity was long described as the degree to which a test measures what it is intended to measure. An updated conceptualization of test validity is that test validity consists of the collection of 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 whether there is test validity evidence to support the intended uses and interpretations of test scores 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 early literacy skills of beginning readers, in the case of the Star Early Literacy Spanish 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 Early Literacy Spanish assessments. These content topics were discussed in detail in "Content and Item Development" and were an integral part of the test items that are the basis of Star Early Literacy Spanish 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 Early Literacy Spanish claims to provide an estimate of the early literacy skills of beginning readers. Therefore, demonstration of Star Early Literacy Spanish's construct validity rests on the evidence that the test provides such estimates. There are a number of ways to demonstrate this.

This section deals with both internal and external evidence of the validity of Star Early Literacy Spanish as an assessment of early literacy skills of beginning readers.

Internal Evidence: Evaluation of Unidimensionality of Star Early Literacy Spanish

Star Early Literacy Spanish is a 27-item computerized-adaptive assessment that measures early literacy skills of beginning readers. Its items are selected adaptively for each student, from a very large bank of early literacy test items, each of which is aligned to one of nine blueprint domains:

- Alphabetic Principle
- Concept of Word
- Visual Discrimination
- Phonemic Awareness
- Phonics
- Structural Analysis
- Vocabulary
- Sentence-Level Comprehension
- Paragraph-Level Comprehension

Star Early Literacy Spanish is an application of item response theory (IRT); each test item's difficulty has been calibrated using the Rasch 1-parameter logistic IRT model. One of the assumptions of the Rasch model is unidimensionality: that a test measures only a single construct such as early literacy skills of beginning readers in the case of Star Early Literacy Spanish. To evaluate whether Star Early Literacy Spanish 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 Pre-K to 3.

To begin, a large sample of student Star Early Literacy Spanish data was assembled. The overall sample consisted of 86,928 student records that took the Star Early Literacy Spanish test in the 2016–2017 or 2017–2018 school years. From that sample, stratified random samples were taken with a sample of 900 students taken for Pre-K, 15,000 students taken for grade K and grade 1, and 10,000 students taken for grades 2 and 3 to yield a total sample of 50,900 students for analysis. These data were the focus of the exploratory and confirmatory factor analyses.

Prior to performing the factor analyses, each student's 27 Star Early Literacy Spanish item responses were divided into subsets of items aligned to each of the blueprint sub-domains. Tests administered in Pre-K aligned to five sub-domains, test administered in grade K aligned with seven sub-domains, tests administered in grade 1 aligned to eight sub-domains, and tests administered in grades 2 and 3 aligned with all nine sub-domains.

For each student, separate Rasch ability estimates (subtest scores) were calculated from each sub-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 1 to 7 items, following the Star Early Literacy Spanish test blueprints, which specify different numbers of items per sub-domain, depending on the student's grade level.

Intercorrelations of the sub-domain-specific Rasch subtest scores were analyzed using exploratory factor analysis (EFA) to evaluate the number of dimensions/factors underlying Star Early Literacy Spanish. Varimax rotation was used. In most cases, the EFA 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) with the results for grade 1 based on the MINEIGEN criterion being an exception. In the analysis of grade 1 data, two eigenvalues greater than 1 were found, but the proportion of variance explained by the second factor was similar to what's seen in the other grades. Figures 2 through 6 show the scree plots and variance explained per factor for Pre-K, K, grade 1, grade 2, and grade 3, respectively.

Figure 2: Scree Plot and Variance Explained by Factor Plot from the Pre-K Exploratory Factor Analysis in Star Early Literacy Spanish

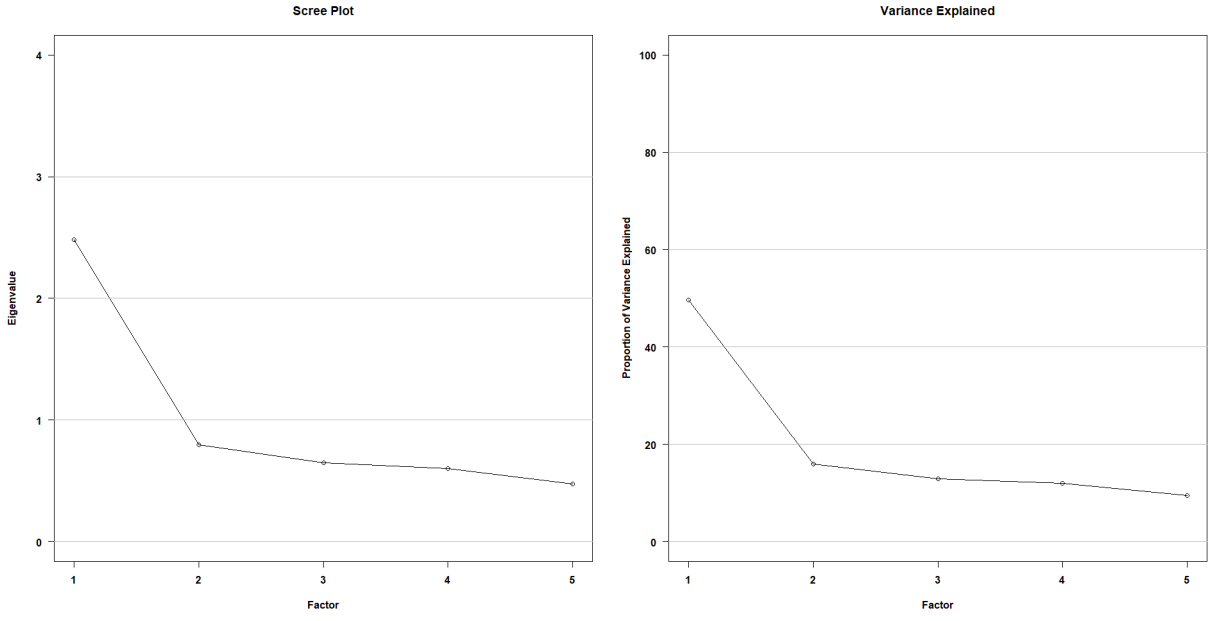


Figure 3: Scree Plot and Variance Explained by Factor Plot from the Kindergarten Exploratory Factor Analysis in Star Early Literacy Spanish

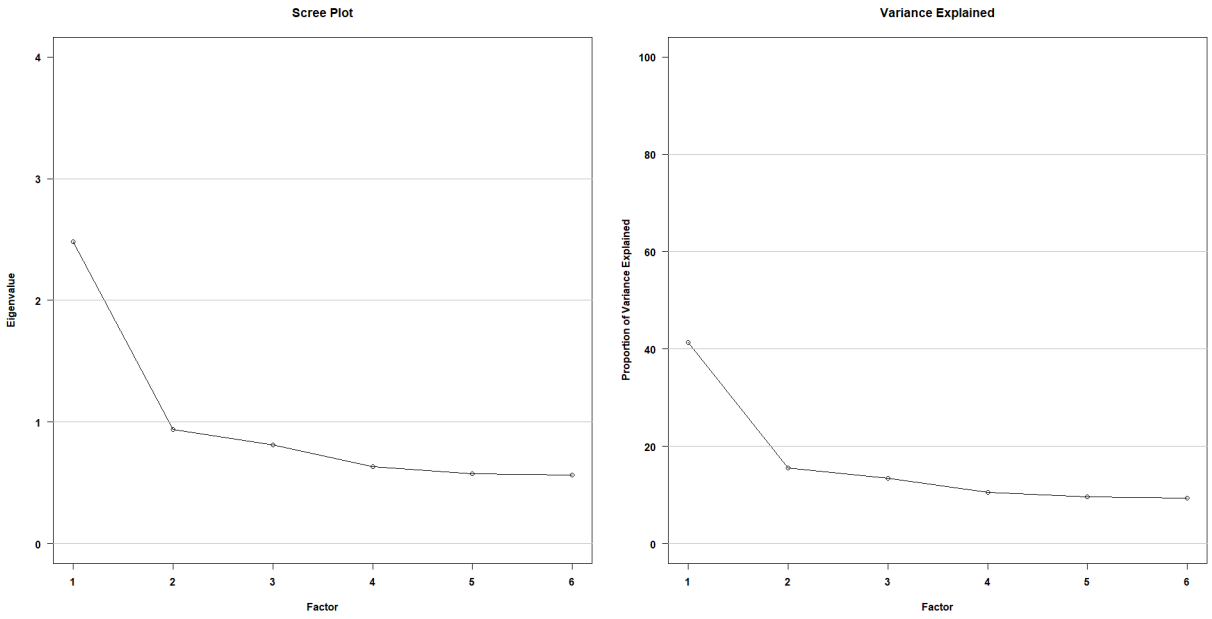


Figure 4: Scree Plot and Variance Explained by Factor Plot from the Grade 1 Exploratory Factor Analysis in Star Early Literacy Spanish

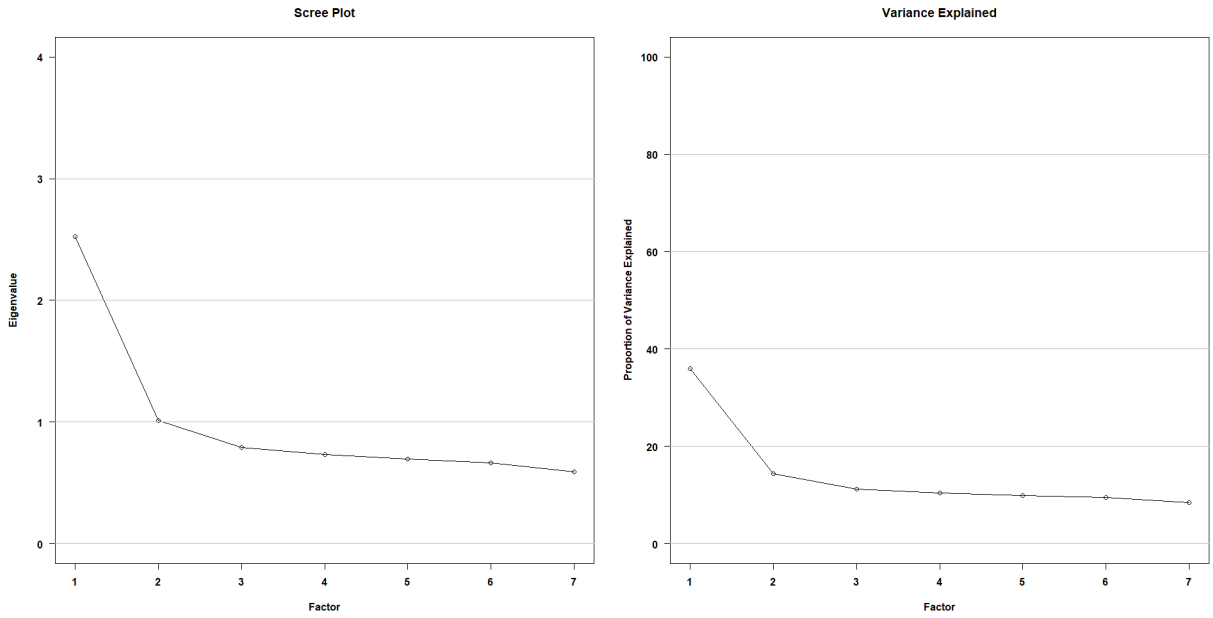


Figure 5: Scree Plot and Variance Explained by Factor Plot from the Grade 2 Exploratory Factor Analysis in Star Early Literacy Spanish

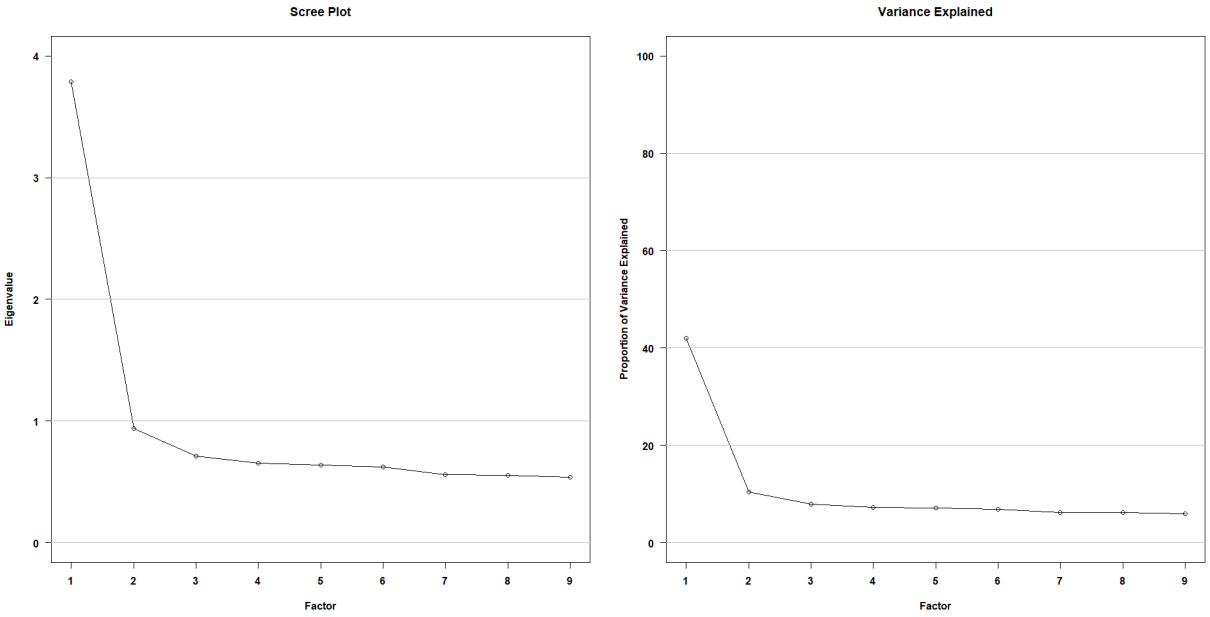
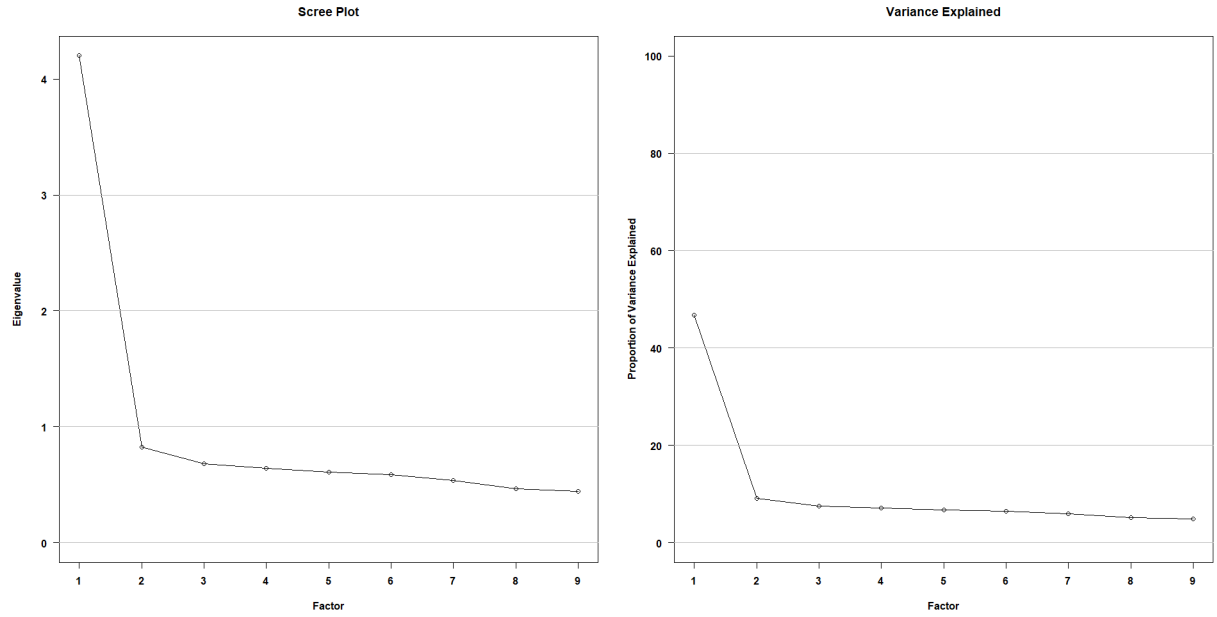
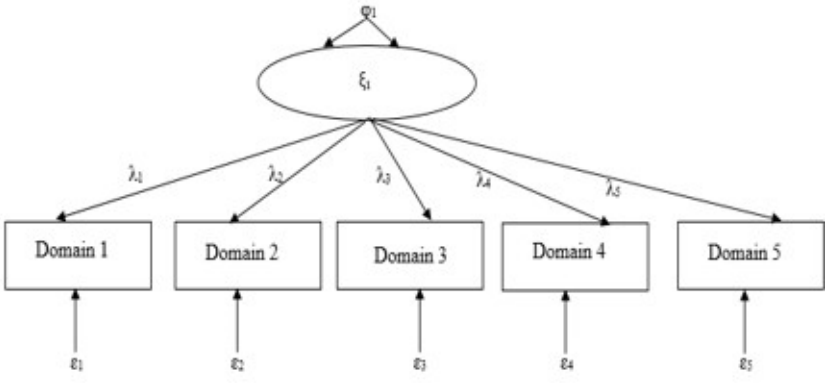


Figure 6: Scree Plot and Variance Explained by Factor Plot from the Grade 3 Exploratory Factor Analysis in Star Early Literacy Spanish



Subsequent to the EFA, 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 based on the number of test domains contained in the blueprint. Figure 7 shows an example of this model assuming five blueprint domains.

Figure 7: Confirmatory Factor Analyses (CFA) in Star Early Literacy Spanish with 5 blueprint domains



The results of the CFA are summarized in Table 9 below. As the table indicates, the sample sizes were large and varied by grade; 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 the Tucker-Lewis index (TLI) are shown; for these indices, values were either 1 or very close to 1 for pre-K and grade 3, indicating strong evidence of a single construct/dimension. The values of the fit indices were a little bit lower for K, grade 1, and grade 2. 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 support a single underlying construct in Star Early Literacy Spanish.

Given that the EFA suggested a potential second factor for grade 1 and that fit indices were a little bit lower for grade K, 1, and 2, two factor models were also fit to data for each grade. Results suggested that there was not consistent interpretation of the two factor models across grades and that different types of content loaded on the two factors depending on the grade level. These results suggest that application of the unidimensional models appears to be appropriate for these tests and that the second factor may represent some degree of noise in the data.

Table 9: Summary of the Goodness-of-Fit of the CFA Models for Star Early Literacy Spanish by Grade

Grade	N	x2	df	CFI	TLI	RMSEA	SRMR
Pre-K	900	10.24	5	0.99	0.99	0.03	0.02
K	15,000	741.25	9	0.95	0.92	0.07	0.03
1	15,000	858.32	14	0.94	0.91	0.06	0.04
2	10,000	1071.08	27	0.95	0.94	0.06	0.30
3	10,000	473.35	27	0.98	0.98	0.04	0.02

The EFA were conducted using the factanal function in R 3.5.1 (R Core Team, 2018), while the CFA was conducted using the lavaan package (Rosseel, 2012) in R.

Types of External Evidence

In an ongoing effort to gather evidence for the validity of Star Early Literacy Spanish scores, continual research on score validity has been undertaken. In addition to original validity data gathered at the time of initial development, a small number of studies have investigated correlations between Star Early Literacy Spanish tests and other external measures. There are generally three types of correlations with external measures that can be explored; concurrent validity estimates, predictive validity estimates, and discriminant validity estimates.

For Star Early Literacy Spanish, concurrent validity is defined as taking a Star Early Literacy Spanish test and another external measure that also assesses reading achievement in Spanish within a 1-month time period. At present, only one concurrent validity study has been conducted since Star Early Literacy Spanish has only been used operationally for a few years.

Predictive validity provides estimates of the extent to which scores on the Star Early Literacy Spanish test predict scores on an external measure of reading achievement in Spanish at a later point in time, operationally defined as more than a month between the Star test (predictor) and the criterion test. No studies of the predictive validity of Star Early Literacy Spanish have yet been conducted. Future studies will explore the predictive validity of Star Early Literacy Spanish as the test continues to be used.

Discriminant validity estimates consist of taking Star Early Literacy Spanish and another external measure that assess another content area besides reading achievement in Spanish (e.g., correlations with a math achievement measure) within a month time period. Typically, the goal is that discriminant

validity estimates are lower than concurrent validity estimates. Only one study has investigated discriminant validity for Star Early Literacy Spanish.

External Evidence: Relationship of Star Early Literacy Spanish Scores to Other Tests of Spanish Reading Achievement

As of the end of 2018, one study has correlated Star Early Literacy Spanish results with two Spanish reading subtest scores for easyCBM. Table 10 provides a summary of those analyses. The easyCBM study took place during the 2015–2016 school year and only involved grade 2. Concurrent validity estimates with easyCBM ranged from 0.67 to 0.72. These coefficients provide solid evidence of the external relationship between the Star Early Literacy Spanish assessments and these other two Spanish reading subtest scores for easyCBM.

Table 10: Correlations Between Star Early Literacy Spanish and Other Spanish Reading Achievement Measures

Test Form	Date	Score	Pre-K		K		1		2		3	
			n	r	n	r	n	r	n	r	n	r
easyCBM												
Spanish word reading	2015–2016	SS	–	–	–	–	–	–	198	0.72	–	–
Spanish sentence reading	2015–2016	SS	–	–	–	–	–	–	198	0.67	–	–

External Evidence: Relationship of Star Early Literacy Spanish to Other Achievement Tests Measuring Math Achievement

As of the end of 2018, one study has examined the relationship of Star Early Literacy Spanish with other achievement tests measuring math content to provide discriminant validity estimates. This study took place in the 2015–2016 school year and looked at the relationship between Star Early Literacy Spanish and the Common Core State Standards Math subtest score with Spanish translations for easyCBM for grades 1 and 2. Table 11 provides a summary of those analyses. Discriminant validity estimates with easyCBM Common Core State Standards Math subtest score ranged from 0.41 to 0.57. These discriminant validity estimates show that the relationship of Star Early Literacy Spanish with these math achievement tests were lower than the convergent validity estimates. These coefficients provide some evidence of expected external relationships between Star Early Literacy Spanish assessments and these other achievement tests measuring content other than Spanish reading achievement.

Table 11: Correlations Between Star Early Literacy Spanish and Other Achievement Tests Measuring Content Other than Spanish Reading Achievement

Test Form	Date	Score	Pre-K		K		1		2		3	
			n	r	n	r	n	r	n	r	n	r
easyCBM												
Common Core State Standard Math Score	2015–2016	SS	–	–	–	–	1,093	0.41	229	0.57	–	–

Summary of Star Early Literacy Spanish Validity Evidence

The validity data presented in this technical documentation includes evidence of Star Early Literacy Spanish's content and construct validity. While the amount of data presented in this technical report is less than the amount of data provided for Star Early Literacy, since the tests has only been in operation for a few years, the data provided was quite positive. The information presented in the "Content and Item Development" chapter supported the content validity of Star Early Literacy Spanish. Exploratory and confirmatory factor analyses provided evidence that Star Early Literacy Spanish measures a unidimensional construct, consistent with the assumption underlying its use of the Rasch 1-parameter logistic item response model. The small number of convergent and discriminant validity estimates indicate that Star Early Literacy Spanish exhibits appropriate moderate to high correlations with other measures of Spanish Reading achievement and that these correlations were higher than correlations with achievement measures in other subjects. Taken together, these data provide support for the claim that Star Early Literacy Spanish is a measure of early literacy skills of beginning readers in Spanish.

Norming

A specific kind of norms, known as test score norms, is described in this chapter. Test score norms are based on distributions of test scores for groups of students at certain points in time. Test score norms are designed to indicate how well a student did on the test in comparison to the groups of students included in the norming samples. Test score norms are not designed to measure changes in ability over time or indicate the amount of growth in ability for individual students. The current version of the Star Early Literacy Spanish test is the first version to have user test score norms.

The 2018 Star Early Spanish Literacy Norms

New US norms for Star Early Literacy Spanish assessments were introduced at the start of the 2018–19 school year. Separate early fall and late spring norms were developed for grades Kindergarten through 3. Due to small sample sizes and the similarity of data for grades 2 and 3, the decision was made to combine these two grades together when creating norms for grades 2 and 3.

The norms introduced in 2018 are based on test scores of K–3 students that took the Star Early Literacy Spanish test during the 2016–2017 or 2017–2018 school years who had complete assessment data. These norms are on the Star Unified Scale.

Students participating in the norming study took assessments between August 1, 2016 and June 30, 2017 or between August 1, 2017 and June 30, 2018. Students took the Star Early Literacy Spanish 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 the Star Early Literacy Spanish tests as they are administered in everyday use.

Sample Characteristics

During the norming period, a total of 86,928 US students in grades K–3 took the current Star Early Literacy Spanish tests. 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 2016–2017 or 2017–2018 school years under normal testing conditions and who had complete assessment data. Data used for the norming analyses consisted of the full sample of students that took the test in either the fall or the spring. If a student took more

than one assessment in the fall the first assessment administered in the fall was included in the norming sample, and if a student took more than one assessment in the spring the last assessment taken was included in the norming sample. Since there is not currently a widely accepted definition of what constitutes a representative national population of US students taking Spanish tests, Renaissance's post-stratification procedure used with Star Reading, Star Early Literacy, and Star Math to make norms nationally representative was not applied to these data. However, data on the percentages in different geographic regions, school enrollments, socioeconomic statuses, school locations, and school types is provided.

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 years was 89,360 students in grades K–3. There were 54,030 students in the fall norming sample and 35,330 students in the spring norming sample. Some students contributed test results in both the fall and spring and in both the 2016–2017 or 2017–2018 school years. These students were counted for each unique assessment in each school year when computing the norming sample size. These students came from schools across the 50 US states and the District of Columbia.

Table 12 provides a breakdown of the number of students participating per grade in the fall and in the spring samples.

Table 12: N Counts per Grade in the Fall and in the Spring Samples

	Fall Sample	Spring Sample
Grade	**N**	**N**
K	22,440	17,920
1	25,430	14,230
2	5,188	2,719
3	972	461
Total	54,030	35,330

Estimates of US student population characteristics for the school included in the norming samples were obtained from the Market Data Retrieval (MDR). The estimates of school-related characteristics were obtained from May 2018 Market Data Retrieval information. The MDR database contains the most recent data on schools, some of which may not be reflected in the NCES data. These data can be directly linked to the assessment data of students included in the norming sample.

Table 13 on page 51 shows the percentages of children in grades K–3 by region, school enrollment, school socioeconomic status, location, and school type nationally, and for the fall and in the spring norming samples. There were some missing data for some students where MDR data could not be linked to the student assessment data. For the fall norming sample 14.18% of the sample was missing MDR data and for the spring norming sample 13.87% was missing MDR data. A brief description of the geographic region, school enrollment, school socioeconomic, location, and school type variables based on MDR is provided 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 was 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%)

School location. Schools were classified into one of four categories based on the school metro code type. The classifications were as follows:

- Rural
- Suburban
- Town
- Urban

School type. Schools were also classified into one of two categories based on whether the school was a public or non-public school.

Table 13 presents the sample characteristic percentages for the MDR variables for the fall and spring norming samples.

Table 13: Sample Characteristics for Fall and Spring Norming Samples

		National Estimates	Fall Norming Sample	Spring Norming Sample
Region	Midwest	20.9%	14.8%	15.1%
	Northeast	19.2%	6.8%	9.9%
	Southeast	24.6%	2.1%	2.0%
	West	35.3%	76.2%	73.0%
School Enrollment	< 200	4.0%	1.2%	2.0%
	200–499	26.8%	22.4%	24.4%
	≥500	69.1%	76.5%	73.6%
District Socioeconomic Status	Low	19.5%	66.8%	69.9%
	Below Median	24.3%	22.3%	19.3%
	Above Median	25.2%	5.9%	5.3%
	High	31.1%	5.0%	5.5%
Location	Rural	14.1%	2.1%	2.8%
	Suburban	42.3%	30.5%	31.5%
	Town	11.7%	4.0%	4.2%
	Urban	31.9%	63.4%	61.6%
School Type	Public	–	98.6%	98.1%
	Non-Public	–	1.4%	1.9%

The norming sample also included students of different genders and ethnicities as well as students with disabilities and English Language Learners. Table 14 provides information on the demographic characteristics of students in the sample. No weighting was done based on 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 14 are based on only those students that had gender and ethnicity information available. Data on students with disabilities and English Language Learners are not provided because many Star assessment users do not enter that information and initial analyses of data in the norming samples suggested that the percentages of students with disabilities and English Languages may underestimate the total percentage of students in these two groups. School type was defined to be either public (including charter schools) or non-public (private, Catholic).

Table 14: Student Demographics and School Information: Samples Percentages for Fall and Spring Norming Samples

			National Estimate	Fall Norming Sample	Spring Norming Sample
Gender	Public	Female	48.60%	49.93%	50.36%
		Male	51.40%	50.07%	49.64%
	Non-Public	Female	–	51.54%	54.63%
		Male	–	48.46%	45.37%
Race/Ethnicity	Public	American Indian	1.00%	6.79%	5.30%
		Asian	5.30%	0.55%	0.58%
		Black	15.50%	2.47%	2.98%
		Hispanic	25.40%	82.53%	81.08%
		White	49.60%	7.66%	10.05%
		Multiple Race[a]	3.20%	–	–
	Non-Public	American Indian	0.50%	0.00%	0.00%
		Asian	6.60%	5.51%	4.39%
		Black	9.10%	0.22%	0.48%
		Hispanic	10.70%	65.42%	53.46%
		White	69.20%	28.85%	41.77%
		Multiple Race[a]	3.90%	–	–

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

Test Administration

All students took the current version of the Star Early Literacy Spanish tests under normal administration procedures. Some students in the norming 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 Early Literacy Spanish test users. Data were from the 2016–2017 or 2017–2018 school years from August to June in each school year. Students' Rasch scores on their first Star Early Literacy Spanish test taken between the first and the second month of the school year based on grade placement were used to compute norms for the fall; students' Rasch scores on the last Star Early Literacy Spanish test taken during the 7th, 8th, or the 9th 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 Rasch scores for each grade.

Table 15 provides descriptive statistics for the norming sample at each grade in the Unified scaled score units.

Table 15: Descriptive Statistics for Scaled Scores by Grade for the Norming Samples on the Unified Scale

	Fall Unified Scaled Scores				Spring Unified Scaled Scores			
Grade	**N**	**Mean**	**Standard Deviation**	**Median**	**N**	**Mean**	**Standard Deviation**	**Median**
K	22,440	692	64	690	17,920	759	70	757
1	25,430	756	67	755	14,230	824	75	829
2	5,188	809	77	817	2,719	850	85	852
3	972	831	86	841	461	861	97	861

Score Definitions

This chapter enumerates the scores reported by Star Early Literacy Spanish, including Scaled Scores, norm-referenced, and criterion-referenced scores.

Types of Test Scores

Star Early Literacy Spanish software provides three broad types of test scores that measure student performance in different ways:

- *Scaled Scores.* Star Early Literacy Spanish 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 Early Literacy Spanish tests to scores on a common scale. Star Early Literacy Spanish 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. In the case that a student gets all items right or wrong, a proprietary Bayesian-modal item response theory estimation method is used. Second, the Rasch ability scores are converted to Scaled Scores. The score scale on which the Scaled Scores are reported is known as the "Unified" score scale.

  Unified Scale Scores

  Renaissance developed a single score scale that applies to all Star assessments: the Unified score scale. That 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 scale scores anchored to important points on the original Enterprise score scales that were developed for Star Math and Star Reading. The end result was a reported score scale that extends from 200 to 1400.

  Star Math, Star Reading, Star Reading Spanish, and Star Math Spanish Unified report scale scores that range from 600 to 1400. Star Early Literacy and Star Early Literacy Spanish 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: test scores 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. The Unified score scale is the only scale used to report results for Star Spanish assessments.

- *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. On the basis of Scaled Scores, students taking Star Early Literacy Spanish are categorized into one of three literacy classifications (see the Literacy Classification section below).
- *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 Early Literacy Spanish software. Both of these scores are based on a comparison of a student's test results to the data collected during the 2018 Star Early Literacy Spanish norming program.

Literacy Classification

Star Early Literacy Spanish score reports include a classification of the student into one of three literacy classifications or reading development stages, based on the Scaled Score. Students with Star Early Literacy Spanish Scaled Scores below 785 are classified as "Emergent Readers," those with scores from 786 through 851 are classified as "Transitional Readers," and those scoring 852 and above are classified as "Probable Readers."

The cut points for these three categories are competency-based. To be classified as a Transitional Reader, a student needs to have mastered specific skills that are represented in the Star Early Literacy Spanish item bank. Similarly, to be classified as a Probable Reader, mastery of higher-level skills must be apparent. Table 21 on page 66 in the Conversion Tables section contains the literacy classifications on the Unified scale for Star Early Literacy Spanish.

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 2.4, this means that the student scored as well on Star Early Literacy Spanish as did the typical student in the fourth month of grade 2. It does not necessarily mean that the student can read independently at a second-grade level, only that he or she obtained a Scaled Score as high as the average second-grade, fourth-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 Early Literacy Spanish Grade Equivalents range from 0.0–3.9. 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 16 indicates how the GE scale corresponds to the various calendar months. For example, if a student obtained a GE of 3.6 on a Star Early Literacy Spanish assessment, this would suggest that the student was performing similarly to the average student in the third grade at the sixth month (March) of the academic year. Because Star Early Literacy Spanish 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 17 on page 61 contains the Scaled Score to GE conversions.

Table 16: 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	0.00	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 higher 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 Star Early Literacy Spanish with Conventional Tests

Because the Star Early Literacy Spanish test adapts to the reading level of the student being tested, Star Early Literacy Spanish 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 first-grade student to obtain a GE score of 3.9 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 third-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 first-grade test.

Star Early Literacy Spanish 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 3.6 indicates that the student's score can be appropriately compared to that of a typical third-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 third-grade reading material).

Percentile Ranks (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 in the national norms group. 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 Early Literacy Spanish as a part of the 2018 Star Early Literacy Spanish norming study. 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 3.7, a Scaled Score of 912 corresponds to a PR of 80, and a Scaled Score of 951 corresponds to a PR of 90. Thus, a difference of 39 Scaled Score points represents a 10-point difference in PR. However, for students at the same 3.7 grade placement, a Scaled Score of 846 corresponds to a PR of 50, and a Scaled Score of 865 corresponds to a PR of 60. While there is now only a 19-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 18 on page 63 in the Conversion Tables chapter contains abridged versions of both the Star Early Literacy Spanish and the Unified Scaled Scores to Percentile Rank conversion table that the Star software uses to look up percentile ranks for each test. The unabridged table includes data for all of the monthly grade placement values from 0.0–3.9 (Kindergarten through grade 3). Because the norming of Star Early Literacy Spanish occurred in the fall and the spring, the first month and last month are empirically based, and the remaining monthly values were estimated by interpolating between the empirical points for the Fall and Spring norms.

Normal Curve Equivalent (NCE) Scores

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 interpolating test scores, calculating averages, and computing correlation coefficients between different tests. For example, in Star 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 19 on page 64 in the Conversion Tables chapter lists the NCEs corresponding to integer PR values and facilitates the conversion of PRs to NCEs. Table 20 on page 65 provides the conversions from NCE to PR. The NCE values are given as a range of scores that convert to the corresponding PR value.

Grade Placement

Star Early Literacy Spanish 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 first-grader in the seventh month with a Scaled Score of 905 would have a PR of 92, while a second-grader in the seventh month with the same Scaled Score would have a PR of 77.

Thus, it is crucial that student records indicate the proper grade and month within grade when students take a Star Early Literacy Spanish 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 using whole numbers, the Star Early Literacy Spanish software automatically adds fractional increments to that grade based on the month of the test. To determine the appropriate increment, Star Early Literacy Spanish 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 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 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 16 on page 56, "Incremental Grade Placements per Month," summarizes the increment values assigned to each month.

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

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

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

Instructions for specifying school years and grade assignments can be found at https://help.renaissance.com/RP and https://help2.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, the Star Early Literacy Spanish 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.

Conversion Tables

Table 17: Star Early Literacy Spanish (SELS) Scaled Score to Grade Equivalent Conversions

Grade Equivalent	Unified Scaled Score	
	Low	High
0	200	697
0.1	698	700
0.2	701	705
0.3	706	710
0.4	711	716
0.5	717	723
0.6	724	730
0.7	731	738
0.8	739	746
0.9	747	754
1	755	762
1.1	763	770
1.2	771	778
1.3	779	786
1.4	787	793
1.5	794	799
1.6	800	806
1.7	807	811
1.8	812	816
1.9	817	821
2	822	823
2.1	824	824
2.2	825	826
2.3	827	827
2.4	828	829
2.5	830	830
2.6	831	832
2.7	833	833

Table 17: Star Early Literacy Spanish (SELS) Scaled Score to Grade Equivalent Conversions (Continued)

Grade Equivalent	Unified Scaled Score	
	Low	High
2.8	834	835
2.9	836	836
3	837	838
3.1	839	839
3.2	840	841
3.3	842	842
3.4	843	844
3.5	845	845
3.6	846	847
3.7	848	848
3.8	849	850
3.9	851	853
3.9+	854	1100

Table 18: Star Early Literacy Spanish (SELS) Unified Scaled Score to Percentile Rank Conversions[a]

	Grade (First Month)					Grade (First Month)					Grade (First Month)					Grade (First Month)			
PR	K	1	2	3	PR	K	1	2	3	PR	K	1	2	3	PR	K	1	2	3
1	200	200	200	200	26	647	713	760	760	51	692	756	821	821	76	739	803	867	867
2	568	625	654	654	27	648	715	764	764	52	694	758	824	824	77	741	806	870	870
3	579	635	665	665	28	650	717	767	767	53	697	760	825	825	78	743	808	872	872
4	582	643	672	672	29	653	718	770	770	54	698	762	827	827	79	745	811	874	874
5	591	651	678	678	30	657	720	773	773	55	699	763	829	829	80	747	813	876	876
6	593	656	685	685	31	658	722	776	776	56	701	765	831	831	81	749	816	878	878
7	601	662	690	690	32	659	724	778	778	57	702	767	833	833	82	752	819	882	882
8	603	665	695	695	33	660	726	781	781	58	704	769	835	835	83	754	821	884	884
9	605	669	699	699	34	662	728	783	783	59	706	771	836	836	84	756	824	887	887
10	612	673	704	704	35	664	729	786	786	60	709	772	838	838	85	759	827	890	890
11	613	676	709	709	36	667	731	789	789	61	711	775	840	840	86	761	830	893	893
12	614	679	713	713	37	669	733	792	792	62	713	776	842	842	87	764	833	897	897
13	618	683	716	716	38	670	735	794	794	63	715	778	844	844	88	767	837	901	901
14	623	686	719	719	39	671	736	796	796	64	716	780	845	845	89	770	840	905	905
15	624	688	723	723	40	672	738	798	798	65	718	781	847	847	90	773	844	908	908
16	625	690	727	727	41	674	740	800	800	66	720	783	849	849	91	776	848	913	913
17	627	693	731	731	42	676	741	803	803	67	722	785	851	851	92	780	852	917	917
18	632	695	735	735	43	679	743	805	805	68	724	787	852	852	93	783	857	921	921
19	634	698	739	739	44	681	744	807	807	69	726	789	854	854	94	788	862	927	927
20	635	700	741	741	45	683	746	809	809	70	728	791	856	856	95	794	867	935	935
21	636	702	745	745	46	684	748	811	811	71	730	793	857	857	96	800	873	943	943
22	638	704	748	748	47	685	749	813	813	72	732	795	859	859	97	809	881	951	951
23	642	706	751	751	48	686	751	816	816	73	734	797	862	862	98	823	892	969	969
24	645	708	754	754	49	688	753	818	818	74	735	799	864	864	99	851	909	994	994
25	646	710	757	757	50	690	755	820	820	75	737	801	866	866					

a. Each entry is the lowest Scaled Score for that grade and percentile.

Table 19: Percentile Rank to Normal Curve Equivalent

PR	NCE	PR	NCE	PR	NCE	PR	NCE
1	1	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
5	15.4	30	39	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	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	62	56.4	87	73.7
13	26.3	38	43.6	63	57	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
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	95	84.6
21	33	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
25	35.8	50	50	75	64.2		

Table 20: 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	4	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	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	80
16.3	18	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	8	40.5	40.9	33	54	54.4	58	69.7	70.4	83
21.1	22.3	9	41	41.5	34	54.5	55	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	45.4	41	58.4	58.9	66	77.6	78.8	91
29.5	30.2	17	45.5	45.9	42	59	59.5	67	78.9	80.2	92
30.3	31	18	46	46.5	43	59.6	60.1	68	80.3	81.7	93
31.1	31.8	19	46.6	47	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	96
33.4	34	22	48.2	48.6	47	62	62.5	72	88.1	91	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	99
35.5	36	25	49.8	50.2	50	63.9	64.5	75			

Table 21: Star Early Literacy Spanish Unified Scale Scores Literacy Classification Score Ranges

Star Early Literacy		
Star Early Literacy Spanish Classification	**Unified Scaled Score Range**	
	Low	**High**
Emergent Reader	200	785
Transitional Reader	786	851
Probable Reader	852	1100

References

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). easyCBM® online progress monitoring assessment system. http://easycbm.com. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

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.

Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wesley.

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.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.

Index

Numerics

A

B

C

D

E

F

G

H

I

K

L

M

N

O

P

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®

 | (800) 338-4204 | www.renaissance.com

Fremont | London | Madison | Minneapolis | New York | Sydney | Toronto | Vancouver | Wisconsin Rapids

181210