



i-Ready Inform de lectura

Curriculum Associates Technical Report | May 2026

Acknowledgements

The creation of this technical manual would not have been possible without the dedication and expertise of the assessment and psychometric teams at Curriculum Associates, as well as the Access and Equity and the Assessment Insights teams.

No part of this document may be modified or further distributed without written permission from Curriculum Associates, LLC. The names of other companies and their products mentioned are the trademarks of their respective owners.

Suggested Citation: Curriculum Associates, LLC. (2026). *i-Ready Inform de lectura Technical Report*. North Billerica, MA: Author.

Table of Contents

Chapter 1. Introduction and Foundations 1

1.1. Chapter Summary..... 1

1.2. Introduction 1

1.2.1. Foundations 1

1.2.2. *i-Ready Inform de lectura* 2

1.2.2.1 About the Assessment..... 2

1.2.2.2 Reported Scores 2

1.2.2.3 Differences between *i-Ready Inform de lectura* and *i-Ready Inform* 2

1.2.3. Target Student Population 3

1.2.3.1 Student Counts and Descriptive Statistics 4

1.3. Structure of *i-Ready Inform de lectura* 4

1.3.1. Key Components 4

1.3.2. Overview of Spanish Reading 6

1.4. Validity Framework 6

1.4.1. Theory of Action 7

1.4.2. The Purpose of Testing..... 10

1.4.3. Assessment Claims..... 10

1.4.3.1 Claims Related to the Interpretation of Scores 10

1.4.3.2 Claims Related to the Intended Uses of the Scores 11

1.4.4. The Validation Process 11

1.4.4.1 Validation 11

1.4.4.2 Types of Validity Evidence 12

1.4.4.3 The Technical Advisory Committee 13

1.5. Organization of the Technical Manual..... 13

1.5.1. Chapter Structure 13

Chapter 2. Test Design and Development 15

2.1. Chapter Summary..... 15

2.2. Introduction	15
2.2.1. Content Classification Taxonomies	15
2.3. The PLD Framework	16
2.3.1. Summary Claims	16
2.3.2. Placement Levels	16
2.3.2.1 Placement-Level Descriptors	17
2.3.3. Anchor Claims and Range PLDs	17
2.3.4. A Principled Approach to PLD Development	18
2.3.4.1 Instructional Relevance and Consideration of Consequences	18
2.3.4.2 Horizontal and Vertical Articulation.....	19
2.3.4.3 The Relationship between PLDs and Spanish Reading Content	20
2.3.4.4 Alignment of Items to the PLD Framework.....	20
2.4. Item Development.....	21
2.4.1. Process Overview	21
2.4.2. Planning and Specifications	22
2.4.2.1 Passage Specifications.....	22
2.4.2.2 Item Specifications.....	23
2.4.2.3 Measures of Cognitive Complexity	24
2.4.2.4 Multimedia Guidelines	25
2.4.3. Item Writing and Editorial Reviews	26
2.4.3.1 Item Writing Training.....	26
2.4.3.2 Editorial Reviews	26
2.4.3.3 Quality Assurance Reviews.....	27
2.4.4. Authenticity Panel Review.....	27
2.4.5. Educator Content Review	27
2.4.6. Psychometric Analyses	28
2.5. Accessibility and Inclusion Framework	28
2.5.1. Systems Approach to Inclusive Design.....	29
2.5.2. Designing for Self-Directed Learners and Teacher Autonomy	29

2.5.3. Designing for Learner Variability	30
2.5.3.1 Broadly Defined Constructs.....	30
2.5.3.2 Multiple Means of Response and Presentation	31
2.5.4. Designing for Safe Use.....	31
2.5.5. Ongoing Evaluation and Iterative Improvement	32
2.6. Embrace All	32
2.6.1. Bias and Sensitivity	33
2.6.2. Development Guidelines	33
Chapter 3. Test Administration.....	34
3.1. Chapter Summary.....	34
3.2. Introduction	34
3.3. Test Administration	34
3.3.1. Before the Test.....	34
3.3.1.1 Role of Test Proctors	35
3.3.1.2 Overview of the Test Session Setup	35
3.3.1.3 Necessary Equipment and Materials	35
3.3.2. During the Test	35
3.3.2.1 Overview of Testing Experience	35
3.3.2.2 Test Expiration Rules	36
3.3.2.3 Monitoring Student Progress and Activity.....	36
3.3.3. After the Test	37
3.3.3.1 Posttest Procedures	37
3.3.3.2 Reviewing Results, Setting Goals, and Basing Instructional Decisions on Results	37
3.3.4. Resources and Support Materials for Administrators and Proctors	37
3.3.5. Testing Windows	38
3.3.6. Administration Time	38
3.4. Accessibility and Accommodations	39
3.5. Test Security and Privacy	39
3.5.1. Supervision	40

3.5.2. Secure Browser	40
3.5.3. Data Transmission and Storage	40
3.5.4. Data Security and Privacy	41
Chapter 4. Computerized Adaptive Testing	42
4.1. Chapter Summary	42
4.2. Introduction	42
4.2.1. Vertical Scale	42
4.2.2. Computerized Adaptive Testing	42
4.3. Test Flows	43
4.3.1. Reading Test Flows	43
4.3.2. Grade-Level Content Restrictions	45
4.3.2.1 Grade-Level Caps	45
4.4. Adaptive Item Selection	46
4.4.1. Starting Items	46
4.4.2. Item and Testlet Selection	46
4.4.3. Domain Termination Logic	47
4.4.4. Test Termination Logic	48
4.5. Proficiency Estimation	48
4.5.1. Starting Proficiency	48
4.5.2. Proficiency Estimation	48
4.5.2.1 Overall Estimates	48
4.5.2.2 Domain Estimates	49
4.5.2.3 Extreme Score Adjustment	49
4.6. Item Bank	49
4.7. CAT Simulation Studies	51
4.7.1. Simulator Structure and Capabilities	52
4.7.2. Simulator Outputs and Usage	52
Chapter 5. Calibration and Scaling	53
5.1. Chapter Summary	53

5.2. Introduction	53
5.3. Key Validation.....	53
5.3.1. Classical Test Theory	53
5.3.2. CTT Indices used for Evaluating Items	54
5.3.2.1 Item Difficulty	54
5.3.2.2 Item distractor response proportion	54
5.3.2.3 Item discrimination.....	54
5.3.2.4 Item distractor-total test correlation.....	55
5.3.3. Item Evaluation	55
5.3.4. Rasch Model.....	55
5.4. Initial Item Bank Creation	56
5.4.1. Vertical Scaling Study	56
5.4.1.1 Scaling Design and Forms.....	56
5.4.1.2 Calibrations.....	57
5.4.1.3 Scaling Constants	57
5.4.1.4 Evaluation.....	58
5.4.2. Standalone Field Testing	59
5.4.2.1 Field Test Forms and Linking Design.....	59
5.4.2.2 Calibration	59
5.5. Embedded Field Testing	59
5.6. Transformation to the Reporting Scale Score.....	60
Chapter 6. Reliability	62
6.1. Chapter Summary.....	62
6.2. Introduction	62
6.3. Test-Level Reliability	63
6.3.1. Marginal Reliability	63
6.3.2. Standard Error of Measurement.....	64
6.4. Reliability for Select Groups	67
6.5. Reliability of Classification.....	71

6.6. Discussion.....	73
Chapter 7. Setting Standards	74
7.1. Chapter Summary.....	74
7.2. Introduction	74
7.2.1. The Basis for Standard Setting	74
7.2.2. Overview of Method	74
7.3. Placement Levels	75
7.3.1. Goals of Setting Placement Levels.....	75
7.3.2. Placement-Level Descriptors	75
7.3.2.1 Policy-Level PLDs.....	75
7.3.2.2 Range PLDs.....	76
7.3.2.3 Item-PLD Alignment	77
7.4. Methods and Procedures	77
7.4.1. Embedded Standard Setting Method	77
7.4.1.1 Empirical Data	78
7.4.1.2 ESS Analyses	78
7.4.1.3 ESS-Inconsistent Item Review and Resolution	79
7.5. Extensions	80
7.5.1. Early Cut Scores	80
7.5.2. Above and Below Grade Cut Scores.....	81
7.5.3. Domain Cut Scores.....	81
7.5.4. Finalizing the Cut Scores	82
Chapter 8. Score Reports.....	85
8.1. Chapter Summary.....	85
8.2. Introduction	85
8.2.1. Levels of Reporting	85
8.3. <i>i-Ready Inform</i> Results Reports (Spanish Reading)	86
8.3.1. Student Level	86
8.3.1.1 Purpose of the Report	86

8.3.1.2 Overall Scale Score.....	86
8.3.1.3 Overall Placements, Domain Scale Scores, and Domain Placements.....	87
8.3.1.4 Lexile® Measures	88
8.3.1.5 Can Dos and Next Steps.....	89
8.3.1.6 Instructional Guidance and Resources.....	90
8.3.2. Class Level.....	90
8.4. For Families Report	92
8.5. Additional Interpretive Guidance	94
Chapter 9. Validity.....	95
9.1. Chapter Summary.....	95
9.2. Introduction	95
9.2.1. Five Sources of Evidence	95
9.2.2. Locating the Evidence	96
9.3. Evidence Based on Test Content	98
9.3.1. Assessment Design and Development.....	98
9.3.2. Task Administration and Accommodations	99
9.3.3. The Adaptive Algorithm and Test Flows	99
9.3.4. Criterion-Referenced Placement Levels.....	100
9.4. Evidence Based on Response Process	100
9.4.1. Measures of Cognitive Complexity and Difficulty	100
9.5. Evidence Based on Internal Structure	100
9.5.1. Measurement Characteristics	100
9.5.2. Reliability.....	101
9.5.3. Classification Accuracy and Consistency	101
9.6. Evidence Based on Relationships to Other Variables.....	102
9.6.1. Linking Studies	102
9.6.2. Correlation Studies.....	102
9.7. Evidence Based on Test Consequences	102
9.8. Discussion.....	103

References 104

List of Tables

Table 1.1. Descriptive Statistics for *i-Ready Inform de lectura*: Early Fall..... 4

Table 2.1. Policy-Level PLDs 17

Table 2.2. General Considerations for Quality Item Development 23

Table 2.3. Anchor Claims Measured by Placement Level..... 25

Table 3.1. *i-Ready Inform de lectura* Recommended Testing Windows 38

Table 3.2. *i-Ready Inform de lectura* Recommended Schedule by Grade 38

Table 4.1. Reading Items by Domain, Grade, and Test Flow 45

Table 4.2. Number of Items by Grade 49

Table 4.3. Number of Items by Domain 50

Table 5.1. Final Vertical Scaling Constants..... 57

Table 5.2. HOSS Values by Grade Level 61

Table 6.1. Marginal Reliability and Standard Error of Measurement 66

Table 6.2. Marginal Reliability and Standard Error of Measurement for Select Groups 68

Table 6.7. Classification Accuracy and Consistency 72

Table 7.1. Policy-Level PLDs 76

Table 7.2. 2025-2026 Cut Scores (Common for Overall and each Domain) 82

Table 7.3. Overall Impact Data 82

Table 7.4. Vocabulary Impact Data..... 83

Table 7.5. Comprehension Impact Data..... 83

Table 7.6. Phonological Awareness Impact Data..... 84

Table 7.7. Phonics Impact Data 84

Table 9.1. Sources of Validity Evidence by Chapter and Claim: Score Interpretation 96

Table 9.2. Sources of Validity Evidence by Chapter and Claim: Intended Score Uses..... 97

List of Figures

Figure 1.1. Theory of Action.	9
Figure 1.2. Organization of Validity Evidence.....	12
Figure 2.1. Illustration of Summary and Anchor Claims: Vocabulary.....	18
Figure 2.2. Overview of <i>i-Ready Inform de lectura</i> Development Process	21
Figure 4.1. Reading Test Flow—Grades K and 1	44
Figure 4.2. Reading Test Flow—Grade 2	44
Figure 4.3. Reading Test Flow—Grades 3–6	45
Figure 4.4. Average Item Information Curve by Grade for Spanish Reading	50
Figure 4.5. Item Difficulty and Student Proficiency Distributions for Spanish Reading.....	51
Figure 5.1. Final cumulative vertical scaling constants.....	58
Figure 6.1. Conditional SEM of <i>i-Ready Inform de lectura</i>	67
Figure 7.1. Embedded Standard Setting Process (Source: Dan Lewis, Creative Measurement Solutions)	78
Figure 8.1. <i>i-Ready Inform</i> Results Report, Student Level: Scale Score and Placements	87
Figure 8.2. <i>i-Ready Inform</i> Results Report, Student Level: Domain Scores and Placements	88
Figure 8.3. <i>i-Ready Inform</i> Results Report, Student Level: Norm Percentile and Lexile (Reading Only)	88
Figure 8.4. Spanish Reading Summary Claim Difficulty Distribution by Grade and Domain	89
Figure 8.5. <i>i-Ready Inform</i> Results Report, Student Level: Can Do Statements, Next Steps, and Additional Resources.....	90
Figure 8.6. <i>i-Ready Inform</i> Results Report, Class Level: Overall Placement Level Distribution	91
Figure 8.7. <i>Inform</i> Results Report, Class Level: Domain Placement-Level Distribution.....	91
Figure 8.8. <i>i-Ready Inform</i> Results Report, Class Level: Dynamic and Customizable Table of Student Results.....	91
Figure 8.9. Partial View of Page 1 of the For Families Report in Spanish Reading.....	93
Figure 8.10. Partial View of Page 2 of the For Families Report	93

Chapter 1. Introduction and Foundations

1.1. Chapter Summary

Chapter 1 outlines the purpose of this technical manual, which is to provide evidence for the validity of *i-Ready Inform*™ de lectura (*i-Ready Inform* Assessment of Spanish Reading). The validity argument outlined in this chapter is structured around specific claims for the use of the assessment and interpretation of its results.

We begin by explaining the conceptual foundations of *i-Ready Inform* de lectura and identifying the target student population. Next, we describe the assessment's structure, detailing its key features and content domains assessed.

Following this, we present the validity framework based on our Theory of Action, which covers the intended purpose of the assessment, its immediate, intermediate, and long-term outcomes, and uses of the scores. We also explain how we collected and organized evidence to support our validity argument. The chapter concludes with an overview of the remainder of the technical manual.

1.2. Introduction

1.2.1. Foundations

This manual presents the technical characteristics of *i-Ready Inform* de lectura, which is part of Curriculum Associates' comprehensive *i-Ready Assessment* suite¹ of assessment and curriculum offerings. These products are designed to work together to support Curriculum Associates' mission to make classrooms better places for teachers and students.

To achieve this mission, our products are designed to (1) enhance students' understanding of their performance and progress throughout the year, (2) equip teachers with data to inform their instructional decisions, and (3) create a common language among students, teachers, and administrators to facilitate discussions about student learning and resource planning.

We believe that valid assessment information and data-driven instruction will lead to increased knowledge, strategic thinking, and engagement among students, teachers, and administrators, ultimately resulting in better student learning outcomes. To ensure the highest quality data from

¹Other components of the *i-Ready Assessment* suite are (a) *i-Ready Inform* for Grades K-12 in an adaptive online format; (b) *i-Ready Growth Monitoring* for Grades K-8 in a short adaptive online format; (c) *i-Ready Standards Mastery* for Grades 2–8 covering Reading and Mathematics in a digital, fixed form; and (d) *i-Ready Literacy Tasks: Benchmarking and Progress Monitoring* in English and Spanish with one-on-one administration, including Passage Reading Fluency (Grades 1–6) and additional literacy tasks (Grades K–3).

i-Ready Inform de lectura, Curriculum Associates uses best practices in test development and psychometric methods based on industry standards, which form the foundation of this technical manual.

1.2.2. *i-Ready Inform de lectura*

1.2.2.1 About the Assessment

i-Ready Inform de lectura is an online, adaptive assessment designed to evaluate student proficiency in Spanish reading for Grades K-6. It is typically administered up to three times per school year (fall, winter, spring). It measures Spanish reading skills in the domains of Phonics, Phonological Awareness, Vocabulary, and Comprehension. The first operational administration of the assessment took place in the 2025-2026 school year.

1.2.2.2 Reported Scores

Both overall and domain-level scores are reported on a vertical scale that ranges from 200 to 500, with increments of one point. Additionally, overall and domain placement levels based on the scale score are reported. Grade-level placements are categorized as Early On Grade, Mid On Grade, and Late On Grade, commonly referred to as Early, Mid, and Late. The Late On Grade placement is not available for Grade 6. At applicable grade levels, students may also be placed below or above their chronological grade level, which is the grade level in which the student is rostered, as reported by their school. An additional placement level, Emerging K, is exclusively for kindergarten and Grade 1. Students in Grades K-4 cannot place more than two grade levels above their chronological grade. Students in Grade 5 can place one grade level above, and students in Grade 6 cannot place above grade level because the assessment goes only to Grade 6. Score reports also include skill-based statements about what a student can likely do and what would be beneficial to work on next. These are referred to as *Can Dos* and *Next Steps*. Chapter 8 provides detailed descriptions and an example of the score reports.

1.2.2.3 Differences between *i-Ready Inform de lectura* and *i-Ready Inform*

i-Ready Inform de lectura is a new assessment. Therefore, some features available for *i-Ready Inform* will be released for *i-Ready Inform de lectura* as operational data become available to support the feature, while other features are not planned. These features include:

- **Personalized Instruction and My Path.** *i-Ready* automatically builds a unique lesson plan for a student consisting of online lessons based on domain-level assessment performance, with a personalized starting point for every student called My Path. Teachers can assign Personalized Instruction lessons manually until the My Path feature is available to assign lessons automatically.
- **Rush Flags.** Operational test data are analyzed to develop test-level yellow rush flags. The goal is to transition to item-level yellow and red rush flags when sufficient operational data are available.

- **User Norms.** Operational user data are used to establish user norms. When sufficient operational data are available and the testing population demographic characteristics are known, it may be possible to provide user norms based on the students taking the assessment.
- **Growth Models.** To develop typical and stretch growth models, data must be collected from the same student across at least two testing occasions. The ability to develop growth measures depends on how many students test within a school year. If enough students test more than once, growth measures may be developed after the first operational administration year.
- **Linking Studies and Projected Proficiency.** Once data are collected from students who take both *i-Ready Inform de lectura* and a state assessment, states interested in conducting a linking study can request one. Linking Studies are then used to calculate Projected Proficiency.
- **Growth Monitoring.** Growth monitoring versions of *i-Ready Inform de lectura* will not be part of the product offering.

As student data increases, we may introduce additional features over time. However, some features may not be offered due to the relatively small number of students taking *i-Ready Inform de lectura*.

1.2.3. Target Student Population

i-Ready Inform de lectura is designed to assess Spanish reading of students in Grades K–6 and is recommended to be administered up to three times per year (fall, winter, and spring) to support Spanish reading instruction in Spanish dual-language, Transitional Bilingual, Spanish Immersion, and English Learner programs.

This assessment was created using original test content developed and written by native Spanish-speaking experts in reading education. Test items were not transadapted or translated from an existing English assessment, resulting in a unique assessment item bank and scale.

The extent to which students with significant cognitive disabilities (i.e., those eligible for their state's alternative assessment of alternate achievement standards [AA-AAS]) can be effectively assessed by, or benefit from, *i-Ready Inform de lectura* largely depends on their individual disability status and proficiencies, as well as the intended use of the assessment for the student. For example, educators might find it beneficial to administer *i-Ready Inform de lectura* to some students with significant cognitive disabilities to gain insights into their foundational reading skills. Currently, the assessment does not have an AA-AAS version aligned with alternate achievement standards, and it has not been evaluated for use with students who have significant cognitive disabilities and are non-speaking or use augmentative and alternative communication (AAC).

1.2.3.1 Student Counts and Descriptive Statistics

Following the first operational year (2025-2026 school year), descriptive information about the students taking the assessment at each grade and for each testing window will be available to include in this technical manual. To provide some early descriptive information, an early pull of data for the fall 2025 administration was analyzed. Data filtering rules were applied to the data to identify the number of valid tests, having only one test per student per testing window. Table 1.1 shows the number of *i-Ready Inform de lectura* tests administered by grade for the first part of the fall testing window (August to mid-October).

Table 1.1. Descriptive Statistics for <i>i-Ready Inform de lectura</i> : Early Fall								
Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	36,480	224	247	264	284	328	268	30.8
1	46,125	247	283	316	342	386	315	42.2
2	42,704	266	317	337	368	405	339	41.4
3	37,389	283	333	364	389	424	361	41.7
4	35,221	300	348	381	404	434	374	41.8
5	30,552	314	362	392	413	446	386	41.1
6	10,887	315	367	397	419	456	392	43.3

1.3. Structure of *i-Ready Inform de lectura*

i-Ready Inform de lectura measures student performance in Spanish reading by assessing student proficiency in the content domains of Phonics, Phonological Awareness, Vocabulary, and Comprehension. The test flow is the order of domains and the number of items administered per domain by grade level. At Grades 3–6, the Phonics domain is administered only if the student’s scale score on Vocabulary and Comprehension is below a certain value. Test flows are discussed in detail in Chapter 4.

1.3.1. Key Components

i-Ready Inform de lectura has several features that support the interpretation and use of the scores. These key features are fundamental to the design and development of the assessment and include:

- **Computer-Adaptive Design.** *i-Ready Inform de lectura* is a computerized-adaptive test (CAT). Its item selection algorithm is applied after each item or reading comprehension testlet to select items with difficulty levels near the student’s estimated proficiency level. By adjusting item selection to student responses while also assessing a wide range of skills, the assessment accurately determines overall and domain proficiency levels.

- **Item Bank.** *i-Ready Inform de lectura* item bank is composed of authentically written items developed by Spanish-speaking reading experts, featuring high-quality content that aligns with meaningful content claims and placement levels. These items are based on both content and cognitive complexity, ensuring they are usable and accessible to the entire target student population. The item bank is monitored to minimize bias or unfairness. It is continuously updated through annual field testing of new items to meet the needs of students across all proficiency levels.
- **Vertical Scale.** The vertical scale for Spanish reading ranges from 200 to 500, enabling score comparisons both within and across grades in aggregate. Individual level interpretations should rely on the criterion-referenced placement levels. The Spanish reading scale differs from the scale used for *i-Ready Inform* for Reading, having its own unique properties and interpretations (see Chapter 1 for descriptive statistics, Chapter 5 for scale transformation constants, and Chapter 6 for information on measurement precision). This design allows for flexibility, enabling students to be placed at, below, or above their grade level.
- **Multiple Administrations.** *i-Ready Inform de lectura* is designed to be administered in the fall (beginning of school year through November 15), winter (November 16 through March 1), and spring (March 2 through end of school year). Seasonal administration allows for proficiency estimates to be updated throughout the school year.
- **Scores and Score Reports.** The scores and score reports for *i-Ready Inform de lectura* are immediately available after the assessment is completed. During the first operational year (2025-2026), report options will be limited, but additional reports will become available in subsequent years. Reports are designed to give educators detailed insights into what Spanish reading skills a student can likely do, propose next steps for skill development, and suggest instructional resources. Students receive an overall scale score representing their performance in Spanish reading, along with domain-level scores reflecting their performance in each domain.
- **Criterion-Referenced Placement Levels.** Scale scores are categorized into placement levels at both the overall and domain level within each grade. For Grades K-5, students can be placed into one of the three on-grade placement levels (Early, Mid, or Late). For Grade 6, on-grade placements are only available for Early or Mid. Students can also be placed above or below grade level, with certain restrictions. The Mid On Grade placement level is regarded as the minimum achievement needed to meet expectations for College and Career Readiness Standards. Chapter 7 details the standard-setting procedures used to establish the thresholds for these placement levels. Standards are periodically reviewed and revised. The same scale score might not correspond to the same placement level across different grades; therefore, criterion-referenced interpretations should be made based on the expectations of the specific grade level.
- **Individualized Learning.** The overall and domain placement levels resulting from an *i-Ready Inform de lectura* administration can guide teachers in assigning Spanish Reading Personalized Instruction lessons to students. These lessons offer students opportunities to improve their

skills between test administrations. Although the Personalized Instruction program is a separate product, it can be used alongside *i-Ready Inform de lectura* to offer more comprehensive support for Spanish Reading development.

1.3.2. Overview of Spanish Reading

The test flow for the *i-Ready Inform de lectura* includes as many as four domains, depending on the grade level:

- Phonics
- Phonological Awareness
- Vocabulary
- Comprehension

Unlike *i-Ready Inform for Reading*, *i-Ready Inform de lectura* does not have a domain for High Frequency Words. The concept of high frequency words is not taught in Spanish reading the same way as it is in English reading instruction. The Comprehension domain includes both literary and informational texts.

i-Ready Inform de lectura test flow varies according to the student's chronological grade level. All students, regardless of their grade level, receive items from the Vocabulary and Comprehension domains. Students in lower chronological grade levels or those performing at lower placement levels may also be given items from the Phonics and Phonological Awareness domains. For a detailed discussion of test flows, refer to Chapter 4.

The Foundational Reading Extension (FRE) feature is not available for *i-Ready Inform de lectura*. However, educators can complement information obtained from *i-Ready Inform de lectura* by administering *i-Ready Spanish Literacy Tasks* to gain additional insight into students' foundational and fluency reading skills in Spanish. This suite of companion tasks can be administered either offline or online, with options for digital data entry and digital scoring.² These tasks for Grades K–6 allow for more targeted understanding of students' reading skills, especially for those needing further evaluation.

1.4. Validity Framework

According to the *Standards for Educational and Psychological Testing* (AERA et al., 2014), validity is defined as “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.” Validity is not determined by a single score or set of scores. Instead, it is established through a structured argument that evaluates whether the intended uses of the

²Access to digital task scoring/administration is not available for Spelling and Encoding task types, which are generally group administered.

assessment are supported or discredited by evidence (Kane, 2006, 2013). Standard 1.0 (AERA et al., 2014) emphasizes the importance of clearly stating each intended interpretation of test scores for a specific use: “Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be provided.” This approach reflects a shift from viewing validity as a fixed property of a test to understanding it as a dynamic process of justification. The goal is to ensure that test scores are interpreted appropriately and used responsibly.

Building a validity argument requires documenting evidence across all phases of the assessment cycle: design, development, administration, and reporting. This process is guided by the validity framework that includes internal, external, and procedural evidence (Kane, 2006, 2013).

The framework begins with a clearly defined theory of action. From this theory, we derive specific assessment claims. These claims form the foundation of the validity argument, which is constructed to evaluate whether the assessment supports its intended interpretations and uses.

Assessment design is the process of determining how to construct an assessment that produces scores suitable for a specific purpose and supports valid interpretation. This process begins with the theory of action, described in the next section. The theory of action outlines the intended outcomes (immediate, intermediate, and long-term) and key steps required to achieve them. Although assessment design work begins the inception of the assessment, it may be revisited and refined over time as the development work progresses.

Assessment development follows the design phase. It involves building the assessment to align with the goals and objectives defined by the theory of action and design specifications.

Both assessment design and development include several critical steps. These steps include planning for the live administration and score reporting, defining the target population (as discussed earlier in this chapter), evaluating available resources and time for administration (see Chapter 3), and determining the level of detail needed for reporting (see Chapters 4 and 5 for measurement details, Chapter 7 for standard setting, and Chapter 8 for score reporting).

Assessment development decisions are guided by the theory of action and design specifications, which together provide a framework for consistent and purposeful development. The theory of action also serves as a benchmark for evaluating the assessment’s effectiveness in achieving its intended outcomes and identifying areas for improvement over time.

1.4.1. Theory of Action

A theory of action is a conceptual framework that describes how a specific intervention, strategy, or set of activities is expected to lead to desired outcomes. It identifies the assumptions, mechanisms, and causal pathways that link inputs to results. By clearly outlining these connections, a theory of

action helps stakeholders understand the rationale for the approach, supports effective planning and implementation, guides evaluation, and promotes continuous improvement.

A theory of action typically includes the following components:

- **Inputs or assumptions:** Resources and beliefs that support the strategy.
- **Activities:** Actions taken to achieve the goals.
- **Outputs:** Immediate results of the activities.
- **Outcomes:** immediate, intermediate, and long-term outcomes or effects.
- **Contextual factors:** Conditions that may influence implementation and success.

At the center of a theory of action is a visual representation—often a flow chart—that shows the relationships among components, activities, and outcomes and their expected long-term impact.

Our theory of action is presented in **Error! Reference source not found.** and shows a series of interconnected components that begin with administering an *i-Ready Inform de lectura* benchmark assessment, then using the assessment results to inform instruction, which leads to improved learning outcomes for students and information to guide instructional programs.

i-Ready Inform de lectura is suitable for use in English Learner (EL) and Spanish dual language/bilingual programs, providing an assessment supported by evidence of reliability and validity to inform literacy instruction in conjunction with other classroom-based information and observations. Intended uses include making student- and classroom-level instructional decisions and informing program-level policy and implementation.

The core of the theory of action is our belief that information about students' proficiency in Spanish reading can inform literacy instruction, so students receive appropriate, targeted instruction or interventions to support development of reading skills necessary, whether in Spanish or English, to attain on-grade level performance and sufficient reading proficiency for academic success.

Various stakeholders (e.g., students, teachers, school and district leaders) play significant roles in the successful use of *i-Ready Inform de lectura* and literacy instruction to address students' reading skill development, and certain activities may be best completed with the use of our other product offerings. This theory of action is embedded in our larger, comprehensive theory of action that drives our curriculum, instruction, and assessment products, as well as professional learning.

i-Ready Inform de lectura

Theory of Action

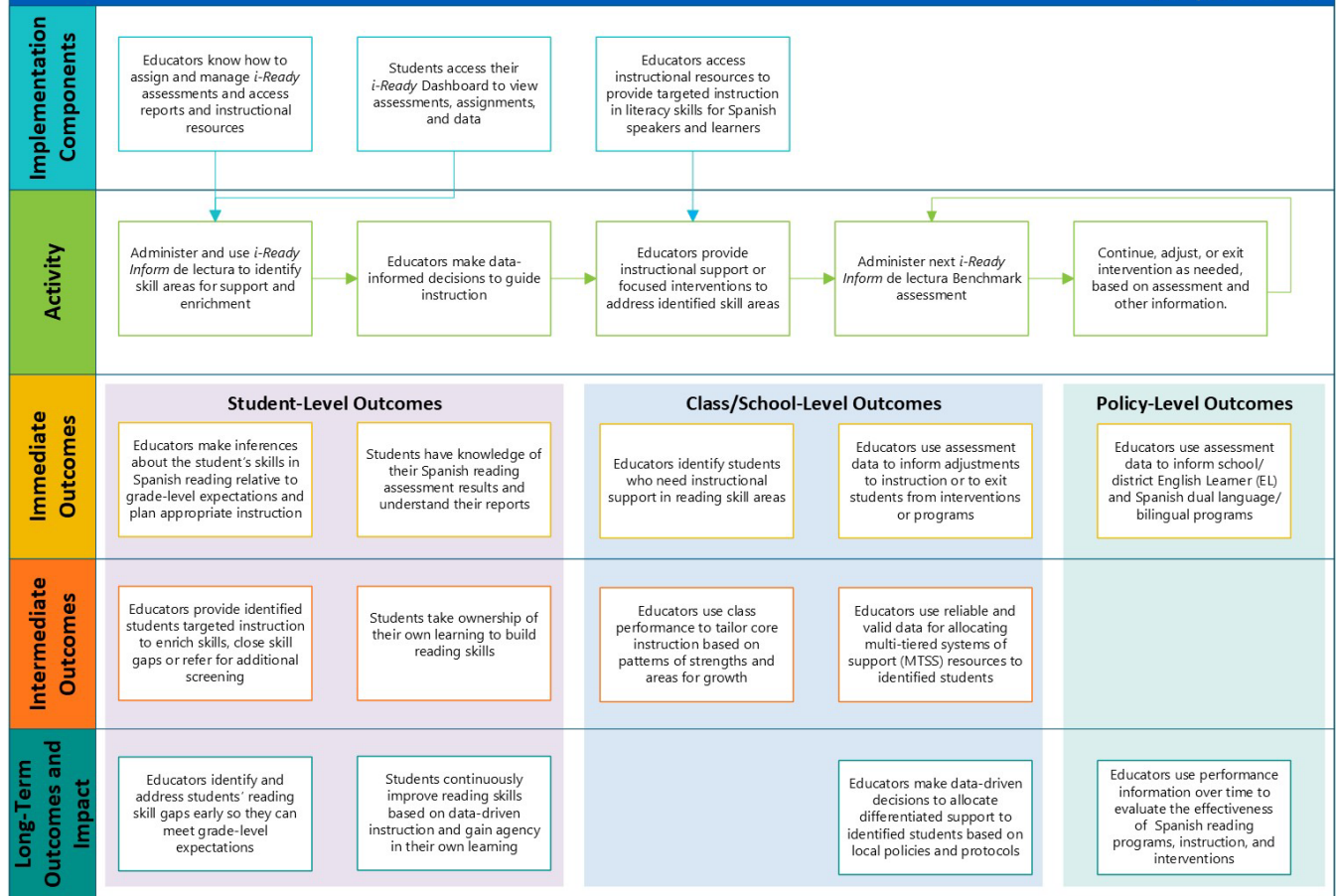


Figure 1.1. Theory of Action. The Theory of Action lays out the rationale for *i-Ready Inform de lectura* and illustrates the relationships among the assessment and the immediate, intermediate, and long-term outcomes of its use.

1.4.2. The Purpose of Testing

The primary purpose of *i-Ready Inform de lectura* is to inform Spanish reading instruction, and, over time, enhance classroom teaching and learning—making it an assessment *for* learning. This purpose, illustrated in the theory of action, is to provide students, teachers, and administrators with precise and timely information. The assessment offers both *descriptive* insights into what students know and can do, and *prescriptive* guidance to support improved student learning. This intended purpose drives the design, development, administrative guidance, and interpretation of the assessment.

1.4.3. Assessment Claims

Assessment claims are broad inferences about how test scores are intended to be interpreted and used. For *i-Ready Inform de lectura*, scores are designed to reflect meaningful information about student performance in relation to expected grade-level knowledge and skills in Spanish Reading.

This section outlines the core assessment claims that anchor the validity framework. These claims are explicit statements describing the intended meaning of scores, assuming the assessment is used as directed. Specifically, *i-Ready Inform de lectura* assessment claims address:

- **Achievement:** How scores reflect student performance relative to the constructs measured.
- **Engagement:** How students, teachers, and other stakeholders are expected to interact with the assessment and its results.
- **Instructional Use:** How assessment results can be validly interpreted and used to support teaching and learning.

1.4.3.1 Claims Related to the Interpretation of Scores

i-Ready Inform de lectura is designed to:

- **Measure** a series of subject-specific, grade-appropriate domains in Spanish Reading.
- **Report** reliable estimates of student proficiency that support comparisons across administrations and grade levels, including:
 - Overall scale scores
 - Overall placement levels
 - Domain scale scores
 - Domain placement levels
- **Contextualize** student proficiency using criterion-referenced placement levels and alignment to applicable state standards.

1.4.3.2 Claims Related to the Intended Uses of the Scores

Validity evidence supports the following uses of scores:

- **Identify and monitor** students who may need additional support, benefit from intervention services, or are ready for more challenging content.
- **Provide actionable feedback** about what students likely can do and what they should work on next.
- **Support differentiated instruction** by enabling teachers to group students based on proficiency levels and specific skill needs, with access to appropriate instructional tools.
- **Assign Personalized Instruction** in Spanish Reading, targeting specific skills where students need additional support.

These uses are grounded in a chain of reasoning that connects assessment claims to design, development, and evaluation processes. The assessment’s layered validity evidence—intentionally designed and transparently communicated—clarifies the intended measurement targets and the constructs to which student inferences can be generalized. This clarity enhances coherence between what the assessment results reveal about student learning and its role within the broader educational context, ultimately supporting meaningful use of results by policymakers, educators, students, and families.

1.4.4. The Validation Process

1.4.4.1 Validation

Validation is the process of developing and supporting a validity argument. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), “Validation logically begins with an explicit statement of the proposed interpretation of the test scores, along with a rationale for the relevance of the interpretation to the proposed use.”

Our validity argument is grounded in the claims presented in the previous section and supported by the rationale outlined in the theory of action. The *Standards* also emphasize that “validation is the joint responsibility of the test developer and the test user.” Curriculum Associates clearly defines how scores from *i-Ready Inform de lectura* should be interpreted and used. It is then the responsibility of users to follow the administration and reporting protocols provided. Using scores as intended—beyond simply administering the assessment—is essential to achieving the greatest impact on student learning.

Validation is not only about the assessment itself; it is about supporting the inferences made based on the assessment results. A strong validity argument requires sufficient evidence to support the intended score interpretations. This chapter outlines the validity principles that guide our approach. Evidence supporting these principles is presented and discussed throughout this document. This

manual concludes with a summary that organizes the validity evidence by type of source and synthesizes the current state of the evidence.

1.4.4.2 Types of Validity Evidence

A theory of action provides the foundation for the evidentiary and validation arguments supporting the assessment, situating the assessment within a broader educational and programmatic context. Our theory of action outlines how *i-Ready Inform de lectura*, along with its reports and instructional resources, work together to improve classroom experiences for both teachers and students.

As an assessment *for* learning, *i-Ready Inform de lectura* generates scores that serve as inputs into other assessment features, such as reports, instructional tools, or supporting products and services that reinforce the theory of action.

Figure 1.2. Organization of Validity Evidence. illustrates how different sets of claims align with key components of the theory of action, based on the framework by Huff, Nichols, & Schneider (in press). Because this manual focuses primarily on the assessment itself, this chapter emphasizes the two rightmost components of the figure (middle and lower levels). The *claims about assessment for given purpose and use* have been discussed in earlier sections.

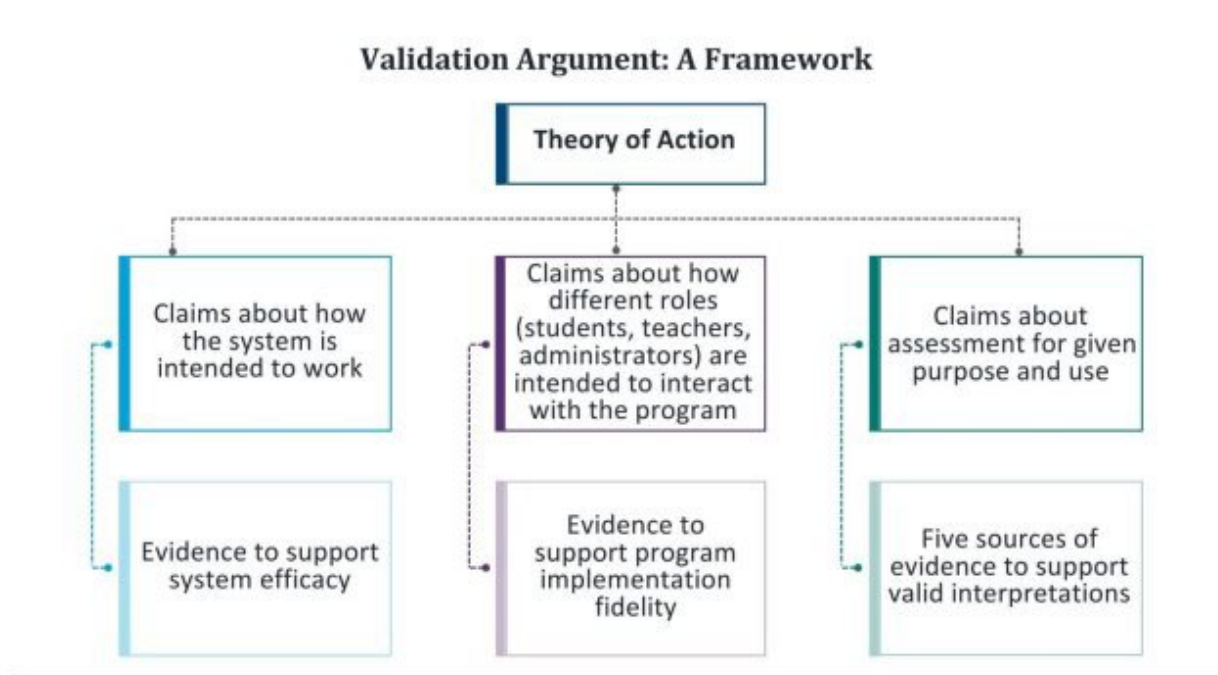


Figure 1.2. Organization of Validity Evidence. Five sources of validity evidence are organized to support the Theory of Action.

There are five sources of evidence used to support valid interpretations:

1. Test content
2. Response processes
3. Internal structure
4. Relationships with other variables
5. Test consequences

Each source aligns with the specific claims made based on the theory of action and provides the evidence needed to support those claims. In Chapter 9, we provide a narrative that integrates these five sources of evidence, providing a framework to support the validity of score interpretations.

1.4.4.3 The Technical Advisory Committee

Curriculum Associates works with a Technical Advisory Committee (TAC) comprised of internationally recognized experts in assessment and psychometrics.³ Key internal teams, including psychometrics, assessment design, and assessment editorial, meet regularly with the TAC to seek guidance on important issues. This ongoing collaboration helps ensure that our approaches to assessment design, content development, and psychometric solutions and analyses meet industry standards and can withstand review by independent experts.

1.5. Organization of the Technical Manual

This technical manual describes the key processes and procedures used by Curriculum Associates to develop, administer, and report scores for *i-Ready Inform de lectura*. Data from operational administrations and validation studies provide evidence to support the overall validity argument.

As we describe the assessment, we explain the design intentions and present supporting evidence aligned with the overarching validity argument.

1.5.1. Chapter Structure

The first five chapters of the technical manual focus on procedural evidence of validity, describing the development and administration of *i-Ready Inform de lectura*.

- Chapter 1 introduces the assessment, including the target student population, test design, theory of action, and assessment claims. These elements are organized around the validity framework. The chapter also outlines the assessment's purpose, intended score uses and interpretations, and an overview of key assessment features.

³As of 2025, the TAC is composed of Dr. Derek Briggs, University of Colorado Boulder, Dr. Andrew Ho, Harvard Graduate School of Education, Dr. Erika Landl, The Center for Assessment, Dr. Richard Leucht, University of North Carolina Greensboro, and Dr. Ye Tong, National Board of Medical Examiners.

- Chapter 2 explains the content development processes, including how the *i-Ready Inform* delectura item pool was created, as well field tested, and monitored over time.
- Chapter 3 describes the administration procedures and accessibility features.
- Chapter 4 describes the computer-adaptive test (CAT) algorithm, including how items are selected and how test flows are managed.
- Chapter 5 presents the psychometric procedures used to build and maintain the vertical score scale. This includes item response theory (IRT) calibration, scaling, equating, score estimation, and quality control procedures.

The next portion of the manual focuses on internal and external validity evidence.

- Chapter 6 provides test reliability metrics and explains how they are used to report and monitor the internal consistency of the assessment. The chapter summarizes work that supports valid interpretation and use of the assessment's criterion-referenced scores and placement levels.
- Chapter 7 describes the standard-setting process used to establish cut scores for the placement levels.
- Chapter 8 outlines the score reports available and the information provided to support understanding of student scores and their aggregations.
- Chapter 9 summarizes the validity evidence presented throughout the report, showing how each piece of evidence supports the assessment claims and aligns with validity framework introduced in Chapter 1.

Chapter 2. Test Design and Development

2.1. Chapter Summary

Chapter 2 outlines Curriculum Associates' approach to content and item development for *i-Ready Inform™* de lectura. It begins with a description of how content is specified, starting with a conceptual framework grounded in best practice and research. The content area of Spanish reading is defined and linked to the development of construct maps, placement levels, and instructionally relevant assessment scores.

Next, the chapter details how assessment items are specified, including how content is sampled, cognitive processes are targeted and measured, and the item types used in *i-Ready Inform* de lectura. The next two sections describe how Curriculum Associates ensures students have full access to the assessment content, including accessibility supports and accommodations.

The final three sections focus on item development and quality assurance: how items are written and reviewed, how they are selected for inclusion using Test Specifications, and how they are field tested and the data evaluated.

2.2. Introduction

The National Research Council (2001) recommended that assessments be grounded in the best available understanding of how students represent knowledge and develop competency within a given domain. A construct refers to the body of knowledge, understanding, skills, and/or abilities that an assessment is designed to measure. Assessments typically measure a hierarchy of constructs, with the highest level (i.e., the coarsest grain size) representing a subject area, and lower levels representing increasingly fine-grained skills.

2.2.1. Content Classification Taxonomies

Curriculum Associates follows best practice and research to define the test constructs. Initially, construct definition involves the use of construct maps, which are then extended to incorporate Placement Levels. This section describes the Placement Level Descriptor (PLD) framework, which is the conceptual framework Curriculum Associates uses to define and operationalize test content.

The elements of the PLD taxonomy explain how key concepts of the framework are related:

- **Grade:** Each item is aligned to a single grade level. Students may receive off-grade items, (see Chapter 4 for additional details).
- **Subject:** The major content area assessed is referred to as the subject. For *i-Ready Inform* de lectura, this is Spanish Reading.

- **Placement level:** Placement Levels form the basis for criterion-referenced score interpretation. These levels were defined and mapped to the score scale during the standard-setting process.
- **Domain:** Within the broader subject of Spanish reading, domains represent finer-grained constructs. The domains are Phonics, Phonological Awareness, Vocabulary, and Comprehension.
- **Summary Claim:** Summary claims describe the overarching skills associated with each grade and domain and may encompass content spanning a wide range of difficulty. They are also aligned to state standards.
- **Anchor Claim:** This is the most granular level of skill categorization within the Placement Level Descriptor (PLD) framework to which items are aligned is the anchor claim. Anchor claims are assigned to a summary claim and a single domain placement level.
- **Item:** Each item is aligned to an anchor claim.

The PLD framework organizes items according to summary claims, domain-level placements, and anchor claims, providing a structured approach to item classification and skill measurement.

2.3. The PLD Framework

2.3.1. Summary Claims

A summary claim describes the knowledge and skills students are expected to acquire within a domain and grade level. It is a broad statement that summarizes the learning objectives and provides an overview of the content students should know and be able to do. Summary claims describe the knowledge and skills in which students should be proficient in the domains of Spanish reading. Several examples from the Grade 4 Vocabulary domain are provided in Figure 2.1 below.

2.3.2. Placement Levels

Placement levels—like performance levels or achievement levels in Grades K–12 assessment programs—define score ranges on a test scale that categorize students’ knowledge, skills, and abilities in meaningful ways. According to the *Standards for Educational and Psychological Testing* (AERA et al., 2014), a performance level is the “[l]abel or brief statement classifying a test-taker’s competency in a particular domain, usually defined by a range of scores on a test.”

For *i-Ready Inform de lectura*, placement levels provide the criterion-referenced context for interpreting student scores. The standard-setting procedures used to define and map these placement levels to the score scale are described in Chapter 7. Current placement level cut scores are available in the document [*i-Ready Inform de lectura Scale Score Placement Tables*](#).

2.3.2.1 Placement-Level Descriptors

A placement-level descriptor (PLD) qualitatively describes what test-takers know and can do at each placement level. Traditionally, such descriptors have been associated with standard setting and score reporting. However, in recent decades, performance standards have also been used to guide test design and development prior to setting cut scores or preparing score reports. PLDs are typically developed within a system of descriptors that vary in granularity—the level of specificity—based on their intended purpose and use (Egan et al., 2012).

Curriculum Associates developed the PLDs within a structured system of interrelated descriptor levels, ranging from broad policy-level descriptors to very specific range descriptors. Policy-level PLDs provide general statements about student performance across all grades and are intended for use in high-level communication with stakeholders such as policymakers and district leaders. In contrast, range PLDs offer more specific descriptions aligned to score ranges and instructional needs, supporting educators in interpreting results and planning instruction. Table 2.1 presents the policy-level PLDs, which apply across all grade levels.

Table 2.1. Policy-Level PLDs	
Placement Level	Description
Below Grade Level	Remediation focused on below-grade level material is recommended to help fill in gaps in students' foundational knowledge. Students in this level are not close to meeting the expectations of College & Career Readiness Standards for their grade level.
Early Grade Level	Students in this level will benefit from on-grade level instruction to help them meet the expectations of College & Career Readiness Standards for their grade level. Students in the early grade level have only partially met these grade-level expectations.
Mid Grade Level	Students in this level will benefit from instruction in late on-grade level topics. These students have met the minimum requirements for the expectations of College & Career Readiness Standards in their grade level.
Late Grade Level	Students in this level will benefit from late on-grade level enrichment and will be ready for instruction focused on topics typically covered in the beginning of the subsequent grade level. Students in the late grade level have successfully met or surpassed the grade-level expectations of College & Career Readiness Standards.
Above Grade Level	Students in this category will benefit from above-grade level instruction. Students in the above grade level have successfully met or surpassed all the expectations of College & Career Readiness Standards for their grade level as well as some expectations from subsequent grade levels.

2.3.3. Anchor Claims and Range PLDs

Following the development of policy-level descriptors, range PLDs were created to provide more detailed descriptions of student performance. These descriptors articulate the specific skills expected at each grade level, the progression of those skills within and across grade levels, and the expected increasing complexity of competency in those skills across placement levels.

Range PLDs are composed of anchor claims—detailed statements of evidence that describe what students are expected to know and do in relation to a summary claim at a given placement level. Each item is mapped to a single anchor claim, and each anchor claims is aligned to a single summary claim and placement level combination. As such, anchor claims serve as the foundation for interpreting student performance at the three on-grade placement levels: Early, Mid, and Late. They define what it means to be proficient in a subject area or domain at each of the three on-grade placement levels.

Figure 2.1 illustrates how these components—summary claims, anchor claims, placement levels, and items—are interrelated in the Placement Level Descriptor framework.

Grade Level Domain	Summary Claim Code	Summary Claim Text	Early Grade Level	Mid Grade Level	Late Grade Level
Grade 4 Vocabulary	ASR.VOC.4.6	Definen el significado del lenguaje figurado, como los símiles, las metáforas, la personificación y las expresiones idiomáticas.	El estudiante determina el significado de lenguaje figurado desconocido, como símiles, metáforas, personificación o frases idiomáticas, basándose en un contexto explícito o fácilmente inferido.	El estudiante determina el significado de lenguaje figurado desconocido, como símiles, metáforas, personificación o frases idiomáticas, basándose en un contexto que requiere un grado moderado de inferencia.	El estudiante distingue entre significados de lenguaje figurado desconocidos que están relacionados, como símiles, metáforas, personificación o frases idiomáticas que están relacionadas entre sí, utilizando un contexto.
		Define the meaning of Grade 4 appropriate figurative language, such as similes, metaphors, personification, and idioms	The student determines the meaning of unfamiliar figurative language, such as similes, metaphors, personification, or idiomatic phrases, based on an explicit or easily inferred context.	The student determines the meaning of unfamiliar figurative language, such as similes, metaphors, personification, or idiomatic phrases, based on a context that requires a moderate degree of inference.	The student distinguishes between related unfamiliar figurative language meanings, such as similes, metaphors, personification, or idiomatic phrases, using context.
	ASR.VOC.4.7	Determinan el significado y el uso de palabras y frases académicas y de dominio específico.	El estudiante determina el significado o el uso de palabras y frases académicas y de dominio específico conocidas, utilizando claves explícitas o fácilmente inferidas del contexto.	El estudiante identifica el significado o el uso de palabras y frases académicas y de dominio específico conocidas, sin un contexto. El estudiante determina el significado o el uso de palabras y frases académicas y de dominio específico desconocidas, utilizando un contexto que requiere de un grado moderado de inferencia.	El estudiante distingue entre los significados o usos de palabras y frases académicas y de dominio específico conocidas que están relacionadas, utilizando el contexto.
		Determine the meaning and use of general academic, conversational, and domain-specific words and phrases appropriate to Grade 4.	The student determines the meaning or use of familiar academic and domain-specific words and phrases, using explicit or easily inferred clues from the context.	The student identifies the meaning or use of familiar academic and domain-specific words and phrases, with no context. The student determines the meaning or use of unfamiliar academic and domain-specific words and phrases, using a context that requires a moderate degree of inference.	The student distinguishes between the meanings or uses of related familiar academic and domain-specific words and phrases, using context.

Figure 2.1. Illustration of Summary and Anchor Claims: Vocabulary

2.3.4. A Principled Approach to PLD Development

2.3.4.1 Instructional Relevance and Consideration of Consequences

Subject Matter Experts (SMEs) for *i-Ready Inform* de lectura are fluent in Spanish with expertise in the assessment of Spanish Language Arts. These SMEs were trained to approach the development of the PLDs through an instructionally grounded lens, with careful consideration of the intended consequences and uses of the assessment. Using a principled approach, the PLD writers began with instructional goals in mind. They considered the full set of content and skills that curriculum, instruction, and assessment should reflect, including the conceptual organization and prioritization of that content.

The content for *i-Ready Inform* de lectura was developed with a deep understanding of the linguistic differences between Spanish and English. For example, Spanish has fewer phonemes than English,

and its syntax and morphology are more regular and transparent. Spanish also features more consistent sound-letter relationships (orthographic transparency) and simpler rime patterns compared to English.

These linguistic characteristics influence how reading skills develop in each language. In Spanish, decoding skills tend to progress more quickly due to the emphasis on syllable segmentation and predictable sound-letter correspondences. In contrast, English phonics instruction often focuses on phoneme blending, pattern recognition, and navigating numerous spelling exceptions. These differences were carefully considered in the design of the assessment to ensure that item content and structure align with the developmental trajectory of reading in Spanish.

From there, the SMEs modeled the domain by developing claims and specifying the evidence needed to support those claims. This work culminated in the creation of a framework that defined how items could be developed within the constraints of the assessment. This framework became the foundation for item specifications.

At this stage, the implications of using the summary and anchor claims were also considered, particularly in terms of how assessment results would inform instruction. As a result, the PLDs were integrated into the principled assessment design approach to serve multiple purposes:

- as a foundation for item specifications, providing a reference for writing content aligned to Early, Mid, and Late placement levels
- as a basis for setting future performance standards (i.e., determining cut scores)
- as a tool for reporting student performance, including *Can Dos and Next Steps*, which are statements directly tied to the PLDs (see Chapter 8)

See Chapter 7 for more information on how placement levels relate to standard setting and cut scores. The relationship between the PLD framework and item development is further explored in the remaining sections in this chapter.

2.3.4.2 Horizontal and Vertical Articulation

During the development of PLDs, SMEs prioritized horizontal alignment of content within each grade level and summary claim. They collaborated with instructional team members to define how skills develop within an anchor claim as students work toward—and sometimes exceed—proficiency in a summary claim.

Using a vertical scale to measure student achievement from one year to the next relies on the vertical articulation of test content across grade levels. Curriculum Associates built vertical articulation into the PLDs as extensions of the construct maps. This articulation ensures that *i-Ready Inform de lectura* can accurately assess student knowledge across grade levels. Additionally, PLDs guide item development to further support the claim that *i-Ready Inform de lectura* accurately assesses student knowledge to gauge improvement over time.

SMEs developed the item pool with a clear focus on grade-to-grade progression. They asked: *Do changes from one grade to the next reflect appropriate developmental and instructional progression?* They designed PLDs to differentiate expectations across grades, increasing the complexity of knowledge and skills in each subsequent year. When summary claims appeared similar across consecutive grades, SMEs added clarifying details to differentiate grade-level expectations. They also ensured that transitions between grades did not include unreasonable leaps in difficulty, maintaining a developmentally appropriate progression throughout the assessment.

2.3.4.3 The Relationship between PLDs and Spanish Reading Content

Curriculum Associates based the PLDs on state Spanish reading standards, including those used in Texas and California. The PLDs represent Spanish reading instructional content delivered in classrooms and define the skills that guide item development.

SMEs began the item writing process by reviewing the PLDs for each grade level. Because the PLDs offer comprehensive coverage of standards and clearly structured horizontal (within-grade) and vertical (grade-to-grade) progressions, they serve as an effective foundation for item design. SMEs used the PLDs to:

- target individual items to a specific anchor claim, intentionally building an item bank that measures distinct aspects of the content
- ensure thorough coverage of all anchor claims by developing strategic item plans that reflect the breadth, depth, and complexity of grade-level standards

2.3.4.4 Alignment of Items to the PLD Framework

Curriculum Associates sources items for *i-Ready Inform de lectura* from two places: the retired K–6 *Assessment of Spanish Reading* and newly developed, field-tested items. To integrate legacy items into the item bank, the team conducted a crosswalk procedure. SMEs aligned each legacy item to the new PLD framework, using a rigorous adjudication process. Meanwhile, item writers developed new items specifically to target anchor claims outlined in the PLDs.

Empirical item data confirmed that item difficulty progressed appropriately across placement levels, reinforcing the vertical articulation built into the PLDs. Based on alignment analyses, the team refined the item pool by disabling or replacing items with difficulty values inconsistent with their anchor claim alignment. Refinements to the item pool included editing and retesting items, refining the wording of an anchor claim, or realigning items to a different anchor claim. Items that did not strongly align with a PLD were disabled and removed from the item bank to maintain the integrity and instructional relevance of the assessment.

2.4. Item Development

2.4.1. Process Overview

Figure 2.2 outlines the comprehensive Item development process for *i-Ready Inform de lectura*. This multi-step process begins with establishing the overall requirements for item development, analyzing content standards, and creating detailed passage and item development plans and specifications to meet the development project’s goal, which may be to create, expand, or refresh the item bank or to address specific gaps in content coverage.

Next, Curriculum Associates trains subject matter experts (SMEs) to write items in accordance with these specifications. As items are drafted, the process continues with editorial and quality assurance reviews and key verification and reconciliation, followed by authenticity panel reviews for fairness and accessibility, educator content reviews, field testing, and empirical evaluation and validation. Each step plays a critical role in ensuring the assessment serves its intended purpose and use and has instructional relevance. The following subsections describe each step in narrative form.

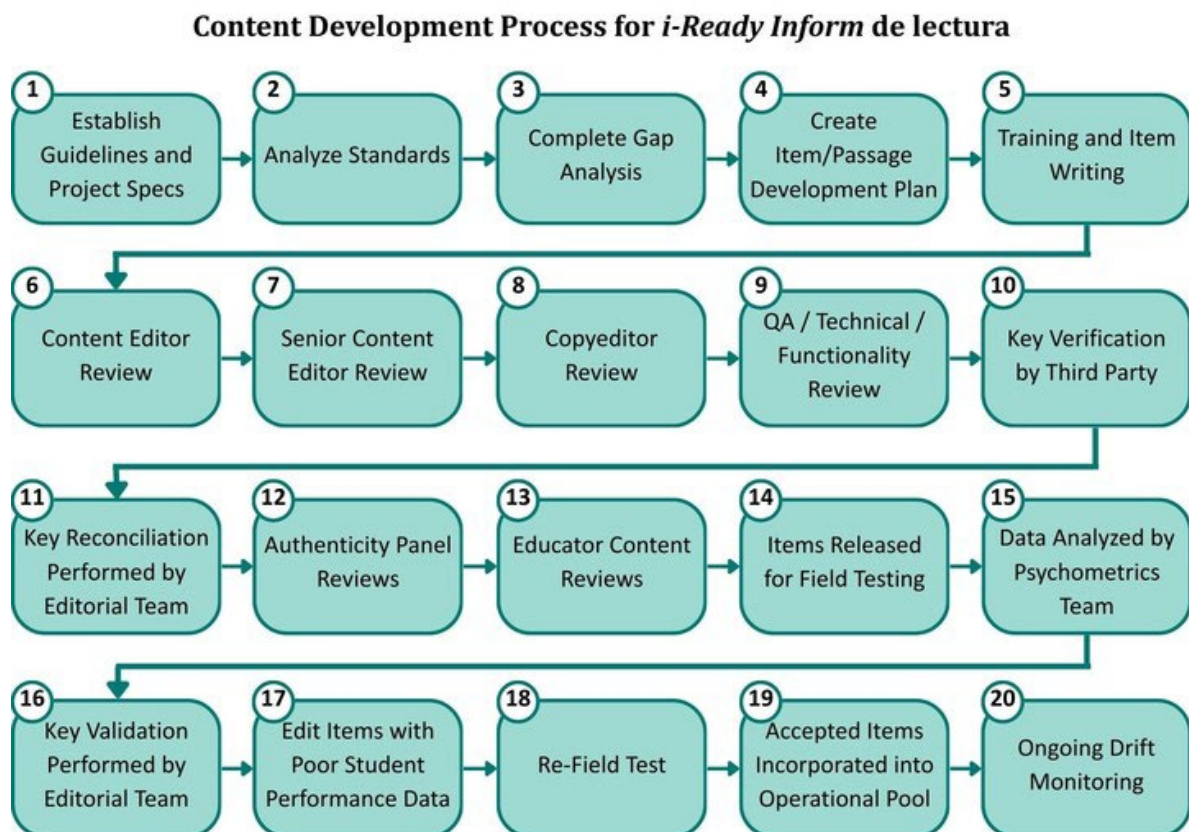


Figure 2.2. Overview of *i-Ready Inform de lectura* Development Process

2.4.2. Planning and Specifications

The item development process begins with overall planning and outlining the **guidelines and specifications** that clarify the development project's purpose and goals and define each step.

First, content editors review which standards to be assessed are already fully covered in the operational item pool and which are partially covered or not covered at all. This *gap analysis* evaluates the robustness of the item bank in relation to anchor claims and standards across the Spanish reading domains and grade levels and informs the more detailed **item and passage development plan**. This plan addresses timelines, staffing, and production goals.

Before item development can begin, the project leads and content editors create detailed requirements for reading passages and items. Detailed passage requirements address the number of passages per grade, distribution by genre, text difficulty and complexity, cultural and linguistic representation, and source—previously published or commissioned for the assessment. Detailed item development plans address the number and type of item per grade and by domain, distribution across standards to be measured and cognitive complexity, and number of items per passage or testlet.

2.4.2.1 Passage Specifications

All Spanish reading passages are either created or sourced from the public domain in accordance with comprehensive specifications. Passage development begins with topic selection, emphasizing constructive and altruistic pursuit (e.g., sports and physical activities, household responsibilities, family support, acts of kindness, and charitable efforts), educational goals and aspirations, environmentally sustainable practices, and culturally and linguistically responsive themes.

Passages are evaluated for word counts, text complexity, genre variety, and relative decodability using the Phonics domain. Vocabulary is reviewed by SMEs with teaching experience at different grade levels to ensure grade level appropriateness. Vocabulary is also reviewed during the Educator Content Review. Because *i-Ready Inform* de lectura is a computer-adaptive test (CAT), and students may encounter passages from various grade levels, explicit grade-level references within passages are avoided (see Chapter 4 for details about the grade-level eligibility of items).

Quantitative measures such as Lexile® levels and word counts are considered in addition to qualitative passage review criteria. SMEs evaluate whether the passage topic and overall conceptual load are grade-level appropriate. They ask questions such as, *Is the topic too mature/complex for the student? Is there too much information to comprehend? Is there too little information provided to comprehend the topic? Can the student (without teacher assistance) understand the information as presented?* Bias and sensitivity are also reviewed using an internal checklist to avoid inappropriate topics.

Passage reviews ensure the accuracy of text information, and introductory text is limited to what is necessary to establish the context. Passages include clearly defined text features such as titles and section headings. Poetry and drama passages may have line numbers, and footnotes are added when

appropriate. Both public domain and commissioned texts are used, including culturally authentic texts originally written in Spanish. To ensure consistency and quality, exemplar passages and review checklists are used throughout the development process.

2.4.2.2 Item Specifications

i-Ready Inform de lectura uses multiple-choice and technology enhanced items (TEIs) such as drag-and-drop, dropdown menu, and text highlight items. Multiple-choice items written for Grades K and 1 have three response choices, while those for Grades 2 and above have four. Although many items are multiple choice, TEIs are a major focus of item development. All items are automatically scored as either correct or incorrect.

Item specifications for *i-Ready Inform de lectura* provide detailed expectations for item development to ensure alignment to the intended anchor claim. Items should not require knowledge beyond the skill level of the anchor claim to answer the item correctly, and SMEs review the items and adjacent anchor claims to confirm proper alignment. In general, item specifications address alignment to content and cognitive complexity, characteristics of item content, quality of the item stem and answer options, and rationales for answer choices. Additional item criteria are outlined in Table 2.2. Art specifications ensure that visual elements and online tools render properly on screen and meet online accessibility guidelines.

Table 2.2. General Considerations for Quality Item Development	
Alignment, DOK, and Rigor	
Item assesses the anchor claim it is aligned to.	
Assigned cognitive complexity measure is accurate.	
Item meets the cognitive demands of both the anchor claim and the complexity designation.	
Item requires comprehension of the passage or multimedia element to be answered.	
Item does not require prior knowledge.	
Item has only one correct response. No arguably correct answers are included.	
Item has less than a 50/50 chance of the student guessing the correct answer.	
Item Content	
Items contain grade-appropriate language, vocabulary, and content.	
Item type is appropriate and effective for the content of the item (especially technology-enhanced items [TEIs]).	
Cueing is avoided between items in a passage set.	
Overlapping/repeated content is avoided between items.	
Items are clear, concise, and complete. • Items include accurate and sufficient content to complete the item. • Items have correct punctuation and are grammatically correct. • Items are free from spelling and typographical errors.	

Table 2.2. General Considerations for Quality Item Development
Items within a passage set cover as much of the text as possible and ask meaningful questions toward understanding the overall content.
Stems
Closed stems are clear, succinct, and follow assessment best practices in construction: • Avoid the use of negatives (e.g., “not,” “none”) in the stem. • Avoid excessive wording (e.g., “of the following”). Only necessary words (1–2 at most) should be emphasized in the stem. • Fiction-based item stems are in present tense.
Answer Options
Avoid cueing within an item (e.g., repeating a word from the stem only in the correct answer).
Distractors are related to the anchor claim being assessed.
All answer choices are plausible and related to or derived from the text without contradicting the text.
Answer options follow assessment best practices in construction: • Avoid outliers. (Answers, or pairs of answers, should all be comparable in language, length, complexity, grammatical form, and have a balance in connotation.) • Do not include pairs of opposites in answer choices. • Avoid throwaway options. • Avoid negative distractors following a negative stem (i.e., double negatives). • Avoid items having absolutes, “none of the above” or “all of the above” as distractors.
Answer-Choice Rationale Criteria
Correct-answer rationales should clearly and concisely explain why an answer choice is correct in relation to the anchor claim, including appropriate references to text content. Each rationale should: • Include relevant language from or about the assessed anchor claim • Be followed by a text-dependent explanation of why the response is correct
Incorrect-answer rationales must describe the mistake, misconception, or misinterpretation the student has likely made as related to the skill in the anchor claim. Each rationale should: • Clearly and concisely explain why a distractor is incorrect in terms of its plausibility within the text • Be followed by text-dependent explanations and/or references that support why the response is incorrect in relation to the assessed anchor claim

2.4.2.3 Measures of Cognitive Complexity

Cognitive complexity—also referred to as cognitive demand—is a key specification in the development of *i-Ready Inform* de lectura items. It reflects the relative complexity of both the content (e.g., simple versus complex text; literal versus figurative language) and the task demands (e.g., level of inferential reasoning required). Each Spanish reading domain applies a domain-specific definition of cognitive complexity and approach to item development.

The skill complexity cognitive demand described in the PLDs drive the complexity and cognitive demand to target in each item. For example, in Vocabulary items, inferencing demands in a PLD reflect increasing sophistication in how students engage with the content as the year progresses. To ensure consistency in how PLDs are written, interpreted, and applied, the team collaborated with experts to finalize the language of the anchor claims.

For example, Table 2.3 shows anchor claims aligned to the summary claim:

Determinan el significado y el uso de palabras y frases académicas y de dominio específico.

Determine the meaning and use of general academic, conversational, and domain-specific words and phrases appropriate to Grade 4.

Anchor claims are articulated across the Grade 4 placement levels, requiring increasingly sophisticated evidence, as indicated by the level of inferencing required.

Table 2.3. Anchor Claims Measured by Placement Level	
Placement	Anchor Claim Text
Early	El estudiante determina el significado o el uso de palabras y frases académicas y de dominio específico conocidas, utilizando claves explícitas o fácilmente inferidas del contexto.
	The student determines the meaning or use of familiar academic and domain-specific words and phrases, using explicit or easily inferred clues from the context.
Mid	El estudiante identifica el significado o el uso de palabras y frases académicas y de dominio específico conocidas, sin un contexto.
	El estudiante determina el significado o el uso de palabras y frases académicas y de dominio específico desconocidas, utilizando un contexto que requiere de un grado moderado de inferencia.
	The student identifies the meaning or use of familiar academic and domain-specific words and phrases, with no context. The student determines the meaning or use of unfamiliar academic and domain-specific words and phrases, using a context that requires a moderate degree of inference.
Late	El estudiante distingue entre los significados o usos de palabras y frases académicas y de dominio específico conocidas que están relacionadas, utilizando el contexto.
	The student distinguishes between the meanings or uses of related familiar academic and domain-specific words and phrases, using context.

Cognitive complexity in *i-Ready Inform* de lectura PLDs is informed by multiple measures of cognitive complexity, including Webb’s Depth of Knowledge (DOK) (1997,1999), Lexile levels of passages, and passage word counts. Cognitive complexity is distinguished from item difficulty, which is determined through item calibrations (see Chapter 5).

2.4.2.4 Multimedia Guidelines

Specifications for art and photographs are attentive to age appropriateness, diverse demographic representations, and ethnic and cultural considerations when people are depicted. Text features—such as charts, graphs, maps, or other text features—are included as needed, along with alternate text for images to ensure accessibility for visually impaired students.

When a video clip, audio clip, or specific image used for multimedia is paired with a passage, introductory text must be provided. This may include the origin of the piece, background information or other relevant context to help students understand what they are about to see or hear. Occasionally a multimedia element appears within a passage. In such cases, a transitional introductory line of text introduces the new element. All Grade K and some Grade 1 passages include audio.

Lexile® is a trademark of MetaMetrics, Inc., and is registered in the United States and abroad. Copyright ©2025 MetaMetrics, Inc. All rights reserved.

2.4.3. Item Writing and Editorial Reviews

2.4.3.1 Item Writing Training

To ensure a coherent testing experience for students, consistency in item design is established throughout the item and test development process. Item writers and editors follow detailed item specifications that govern the structure and format of each item type: item prompts, answer choices, art, and audio and visual presentation. Specifications also address stimulus features and the characteristics, rationale, and limitations of distractors.

Writers participate in in-depth training on the item specifications for each domain and grade level. Training includes instruction on writing items at appropriate levels of cognitive complexity based on depth of knowledge (DOK) (Webb, 1997, 1999) and aspects of rigor (AOR) (Achieve, 2019), as relevant to Spanish reading. To support the writers, exemplar items illustrate each domain and item type, including specific guidance on content limits and accessibility considerations. They then practice writing items that are clearly aligned to the targeted anchor claims and placement level and receive feedback before completing assignments.

2.4.3.2 Editorial Reviews

Curriculum Associates content editors review the writers' item and/or passage submission against a checklist based on the specifications to verify that all requirements are met. In some cases, a simple fix by the editor is all that is needed. In other cases, the assignment is returned to the writer with instructions on how to make this and future items or passages acceptable for use.

Content editors complete their review of the item or passage and notifies the senior content editor. The senior content editor ensures that the specifications are still fully met if the content editor made any edits. The senior editor addresses any concerns with the content editor before moving the item or passage to the next review step.

Copy editors specialize in copyediting items and passages for grammar, usage, and adherence to the style guide specifications. Any changes are reviewed by content editors or senior editors to verify that copyedits do not affect the content of the items.

2.4.3.3 Quality Assurance Reviews

Quality assurance reviews verify that test-takers can respond to the items without any technical difficulties. This review covers sound, video, art, virtual tools, and navigation, among many checks.

Before finalizing items, **key verification** is performed to ensure scoring accuracy. To do this, three people who have not worked on the project take each item and record their answers, which are compared to the key recorded in the content authoring system. Any differences among answers are reconciled by the content editorial team by performing a thorough review of the item. Based on this verification process, possible outcomes are to keep the answer key as is, correct the answer key, or edit the item so there is only one clearly correct answer and repeat the key verification process.

2.4.4. Authenticity Panel Review

The next stage in the item development process is the authenticity panel review, which is one way we promote authentic cultural representations in our products. Panelists volunteer their time and expertise to help make our products more accurate and respectful of the varied cultural experiences of teachers and students. We define expertise broadly—it may come from personal experience, academic research, work with students, and other relevant backgrounds.

These reviews are guided by detailed, content-specific checklists and consider all rings of culture—ethnicity, religion, socioeconomic status, disability, and more—to ensure our products reflect the full spectrum of human identity.

2.4.5. Educator Content Review

Current classroom teachers are recruited to participate in the Educator Content Review to evaluate the passages and items created for *i-Ready Inform de lectura*. Curriculum Associates supports these reviewers by providing all necessary materials—including passages and items and a tracking tool for documenting feedback.

Educators are asked to evaluate several key areas:

- alignment to standards and PLDs
- grade-level appropriateness of passage and item content
- vocabulary suitability
- potential bias or cueing issues

Reviewers' feedback is discussed collaboratively to ensure clarity and shared understanding. Following these discussions, editors reconcile the feedback and implement recommended edits to refine and improve passages and items.

2.4.6. Psychometric Analyses

Before newly written and reviewed items may be added to the operational item bank, they must undergo field testing. Field testing serves two primary purposes:

- It allows us to conduct classical statistical analyses to ensure that the item is functioning as expected and is keyed correctly.
- It allows us to estimate the item difficulty parameter using the Item Response Theory (IRT) model upon which the score scale is built.

Classical item statistics help us identify any item flaws or unexpected behavior. IRT item difficulty parameters are used in the adaptive item selection algorithm and in determining the final score and associated measurement error (see Chapter 5 for more details).

Items that do not perform as expected are flagged and reviewed by SMEs using a **key validation** process. Sometimes the review identifies the need for revisions to an item, which requires that it be field tested again. Alternatively, items may be accepted as is or removed from the item bank.

Periodically, operational items are re-field tested to verify that they continue to function according to the parameters on file. This is known as item drift analysis.

2.5. Accessibility and Inclusion Framework

The Accessibility and Inclusion Framework articulates a comprehensive vision for an inclusive assessment and integrating accessibility into every component of the assessment. Recognizing that both technology and student needs evolve, the framework includes an evaluation and continuous improvement plan as part of the validation process.

Accessibility and inclusion are not treated as isolated goals or as add-ons. Instead, they are embedded throughout the *i-Ready Inform* de lectura Theory of Action (see Chapter 1), with the intent of having a causal impact on outcomes for all learners. Specifically, the framework is designed to support the following intermediate and long-term outcomes, leading to the goal of increased reading achievement:

- students gain agency in their own learning
- educators make data-driven decisions to provide instruction
- educators evaluate the effectiveness of Spanish reading programs, instruction, and intervention

The framework also outlines how the impact of design considerations will be evaluated through studies that supplement the validity plan (described below). This ensures that accessibility and inclusion remain central to our iterative product improvement and development—not as afterthoughts, but as foundational design principles.

2.5.1. Systems Approach to Inclusive Design

Our approach to accessibility and inclusion recognizes the importance of inclusive design throughout all phases of testing—from item development to test score use. Recognizing that assessments like *i-Ready* are used within broader educational systems, we aim for consistency in accessibility features and tools, even when test purposes and policies differ. Our systems-based approach is grounded in universal design (Laitusis & Karvonen, 2024), which centers on three core principles that impact students, teachers, parents, caregivers, and administrators. These principles require that assessment systems: (1) meet basic human needs for autonomy, competence, and relatedness; (2) maximize flexibility for learner variability; and (3) promote equity and mitigate unintended consequences. This framework also integrates federal accessibility requirements, such as the W3C Web Content Accessibility Guidelines (WCAG) (World Wide Web Consortium, 2025), and draws on Universal Design for Learning (UDL) (CAST, 2024) to support diverse learner needs.

2.5.2. Designing for Self-Directed Learners and Teacher Autonomy

The *i-Ready Inform* de lectura Theory of Action identifies students taking ownership of their own learning to build reading skills as a key intermediate outcome. To support this goal, our design framework prioritizes features that enhance student motivation, engagement, and measurement of student proficiency across the full achievement continuum. We also acknowledge the impact of assessments on educator motivation and engagement. Accordingly, both student- and teacher-facing elements of *i-Ready* are designed to embed the three psychological nutrients—autonomy, competence, and relatedness—that foster meaningful goal setting and sustained motivation (Ryan & Deci, 2018). Our design team uses guiding questions to ensure these are thoughtfully integrated within *i-Ready* while maintaining the standardization necessary for score comparability.

The following guiding questions support inclusive design by addressing the psychological needs of autonomy, competence, and relatedness.

Autonomy

- Do students have access to tools, accommodations, and assistive technology and the ability to independently choose when to use them?
- Do teachers have flexibility to select accommodations that yield the most instructionally relevant information?
- Do teachers and administrators have choice in when, where, and how often assessments are administered? For example, are test days and times flexible enough to avoid conflicts with religious holidays, cultural events, or student health needs?

Competence

- Does test content elicit evidence of skills within a student's zone of proximal developments (e.g., with adaptive test design, growth scores)?

- Do accommodations and universal tools remove construct-irrelevant barriers, allowing all students to show what they know and can do?
- Do score reports provide understandable information that helps students and teachers set achievable instructional goals?
- Do teacher score reports offer actionable insights to guide instructional next steps?

Relatedness

- When possible, are accessibility tools and accommodations embedded in the testing platform to promote student autonomy and reduce stigmatization?
- Do test items reflect a wide range of lived experiences and ways of knowing, reflecting diversity in culture, language, and abilities?
- Are alternative pathways provided when certain constructs are not appropriate for a student due to disability or English language proficiency (e.g., adaptive pathways for foundational reading skills in students with hearing impairments)?

2.5.3. Designing for Learner Variability

Designing for learner variability is a core component of the *i-Ready* framework, influencing test design, item development, and test administration. Our CAT design and vertical scale acknowledge the diversity of students across all levels of academic achievement. Adaptive testing presents a unique challenge: every item in the pool must be accessible to students with a wide range of abilities.

Our approach begins with evaluating each domain for appropriateness. When a domain is appropriate for a student's unique abilities, we ensure that every item for the domain is accessible. In cases where a domain is not appropriate, such as audio-dependent phonics items for students with hearing impairments, we provide a "skip foundational domain" feature.

This commitment to accessibility is an ongoing process that includes broadly defined constructs, inclusive item development, and accessible delivery platforms. Several key areas of consideration are described in the sections that follow.

2.5.3.1 Broadly Defined Constructs

Each construct must be defined in a way that allows students with diverse needs to demonstrate their skills meaningfully. For example, foundational reading skills such as decoding should be defined to allow students with hearing impairments to show mastery of foundational word recognition skills. Key questions that guide this work include:

- Do construct definition teams include individuals with expertise in disability-specific instructional approaches for the construct being measured?
- Are alternate pathways to mastery considered when developing progressions of learning?

- Do teachers have autonomy to select accommodations that may alter the construct but still provide valuable insight into the root cause of a student's learning challenges or help set achievable learning goals?

2.5.3.2 Multiple Means of Response and Presentation

While some adaptive tests provide a limited item pool for students with sensory disabilities, the *i-Ready* development team is committed to ensuring that all items within appropriate domains are accessible to every student. Our ongoing work aligns with WCAG 2.2 Level AA (Worldwide Web Consortium, 2025) criteria and incorporates the pedagogical and usability needs of elementary learners.

We provide three levels of supports: universal supports, designated supports, and accommodations. Additional details and educator guidance are available in Chapter 3. Our item development and delivery teams consider the following questions:

- Does the item development team include individuals with lived experiences and knowledge of disability-specific approaches to instruction and assessment?
- What evidence of proficiency will the item elicit when presented in audio, sign language, tactile, or visual formats?
- What evidence of proficiency will the item elicit when responded to in audio (including sign language), tactile, or visual formats?
- How can the item type meet WCAG guidelines and be perceivable, operable, understandable, and sufficiently robust across current and future technologies?
- Does the item pool reflect a broad range of lived experiences?

2.5.4. Designing for Safe Use

Designing for safe use is the final core principle of the Universal Design of Assessment Systems. Operationalizing this principle involves professional learning to build assessment literacy among score users, the development of a theory of action, ongoing validity studies, and safeguards against score misuse. While these principles apply to all test-takers, special attention is needed for students with disabilities and multilingual learners.

By integrating accessibility and inclusion into each component of the *i-Ready Inform de lectura* Theory of Action, we aim to serve the diverse needs of all students and foster an inclusive learning environment where every learner can thrive. This section outlines how the core design principles apply across the assessment system.

Because student demographic information is protected under the Family Educational Rights and Privacy Act (FERPA), identifying schools and districts willing to participate in research is essential in gathering reliability and validity evidence to evaluate the *i-Ready Inform de lectura* Theory of Action

for all student groups. Once partner schools are identified, both quantitative and qualitative research should be conducted to explore key questions related to safe use, such as:

- Do special education and general education teachers report similar utility of *i-Ready* scores for:
 - Measuring student learning
 - Increasing self-directed learning
 - Understanding the root cause of learning challenges
 - Interpreting score reports and using them effectively
- Do special education and general education students report similar abilities to:
 - Access test questions independently
 - Answer test questions independently
 - Demonstrate what they know and can do (with appropriate accommodations)
 - Understand score reports and use them meaningfully

2.5.5. Ongoing Evaluation and Iterative Improvement

Validation is an iterative and ongoing process. As new assistive technologies emerge and additional student needs are identified, *i-Ready*'s accessibility features continue to evolve. We are committed to improving accessibility and inclusivity through teacher and student feedback gathered through self-report surveys, cognitive labs, and systematic analyses of customer feedback.

In addition, we actively seek partner schools to collaborate on secondary analyses of user data. These studies may include differential item functioning (DIF) by accommodation, language, race, gender, and/or disability group, as well as user data from teachers in various instructional settings—general education inclusion, specialized instruction classrooms, and specialized schools. As technology advances and our network of research partner schools grows, we will evaluate how iterative design changes impact academic growth across student groups.

2.6. Embrace All

Curriculum Associates' mission is to make classrooms better places for all teachers and students. We believe engaging materials paired with strong pedagogy foster academic excellence and lead to high achievement for all students—regardless of cultural identity, linguistic background, economic status or circumstance, and disability. We also believe that connection creates impact, and connection requires relevance. Relevant instruction happens in classrooms where students feel safe, accepted, and seen.

Curriculum Associates is committed to:

- celebrating students' identities and helping teachers use students' experiences to shape content and instruction
- supporting all students in becoming independent learners and engaged global citizens

These values guided the development of *i-Ready Inform de lectura*.

2.6.1. Bias and Sensitivity

Bias in assessment occurs when a construct-irrelevant characteristic leads to differential performance for students with the same proficiency but from different groups. Curriculum Associates works to avoid bias by developing culturally and linguistically responsive content. Bias may occur through over- or underrepresentation of any given student group.

Our Spanish reading content, originally written in Spanish, reflects diverse student groups—across ethnicity, national origin, race, region/location, age, gender, socioeconomic class, religion, and disability—as realistically as possible. We strive to represent multiple viewpoints, remain true to the culture of origin, and reflect everyday experiences of people from a variety of cultures and communities.

2.6.2. Development Guidelines

Curriculum Associates' product development guidelines outline key steps for developing content for *i-Ready Inform de lectura*:

- Collaborate to develop and document a formal plan for cultural relevance and responsiveness during the discovery and definition phases.
- Build a well-rounded development team—including writers, editors, designers, artists, vendors, reviewers, and program authors, and SMEs—from diverse communities, including individuals with disabilities and English Learners.
- Consider cultural relevance and responsiveness when developing the visual strategy and image plan. Images are a powerful form of communication and should reflect varied representation. Background settings (including people and context) should be specified, discussed, and tracked when only the foreground is defined.

In summary, Curriculum Associates is committed to creating inclusive, culturally responsive assessments that reflect the diverse identities and experiences of all students. Through intentional design, bias mitigation, and development practices, *i-Ready Inform de lectura* supports equitable learning by honoring student backgrounds, ensuring accessibility, and promoting relevance in both content and presentation.

Chapter 3. Test Administration

3.1. Chapter Summary

This chapter outlines the procedures for administering *i-Ready Inform™* de lectura, including:

- testing windows
- administration protocols before, during, and after testing
- test expiration rules
- accessibility and accommodations.

Accessibility features and accommodations are designed to expand access for as many students as possible, including students with disabilities and multilingual learners. Accessible and consistent administration ensure score interpretation and use as intended.

This chapter also addresses test security protocols and monitoring strategies that promote consistent administration practices and create appropriate conditions for students to demonstrate their proficiency. These practices help support valid interpretations and uses of scores.

Finally, system security protocols are presented, which protect student privacy and safeguard student data.

3.2. Introduction

Detailed and thorough test administration protocols are essential to maintaining the integrity, reliability, and validity of any assessment system. Administering *i-Ready Inform* de lectura is a multiphase process that requires careful planning, attentive implementation, and thoughtful follow-up.

The *i-Ready* delivery platform plays an essential part in this process. Developed according to best practices in assessment design and administration, the platform provides a user-friendly interface for students and robust support for educators. It ensures effective test delivery and offers comprehensive resources to guide each phase of administration.

3.3. Test Administration

3.3.1. Before the Test

i-Ready Inform de lectura is typically administered three times per year: fall, winter, and spring. However, educators may choose to administer it more frequently based on instructional needs.

Administration is managed at the school and classroom levels and delivered through a secure online system. Key roles in the process include school administrators and teacher proctors, who ensure consistent and secure testing conditions.

3.3.1.1 Role of Test Proctors

Proctors play an important role in the effective administration of *i-Ready Inform de lectura*. They are responsible for creating a supportive testing environment, ensuring all necessary materials are available, and guiding students throughout the assessment experience.

Proctors' responsibilities extend beyond the day of the test and include pretest preparations and posttest procedures to help ensure consistent administration processes.

3.3.1.2 Overview of the Test Session Setup

Proper setup of a test session in the *i-Ready Connect*™ platform is essential for smooth administration. Key steps include:

- ensuring necessary equipment is available and functioning
- confirming the test is assigned to the correct students
- preparing the testing environment to minimize distractions
- verifying student readiness to begin the assessment

Thorough preparation helps ensure a consistent and secure testing experience for all students.

3.3.1.3 Necessary Equipment and Materials

To administer *i-Ready Inform de lectura* via the *i-Ready Connect* platform, educators must ensure access to compatible devices⁴ with reliable internet connections and headphones.

In addition, any physical materials required for non-embedded accessibility supports and accommodations (e.g., scratch paper and pencil, dictionary, etc.) should be arranged prior to the start of testing.

3.3.2. During the Test

3.3.2.1 Overview of Testing Experience

i-Ready Inform de lectura follows a structured set of protocols to ensure consistency across test sessions. Key components include:

⁴Information regarding system requirements and supported devices can be found here: https://cdn.i-ready.com/instruction/content/system-check/iReady_System_Requirements.pdf.

- initial login
- overview of test rules
- practice items
- introduction and response to assessment items.

To support student engagement and performance, the assessment includes intermittent breaks (e.g., Brain Breaks) to encourage student engagement and sustain motivation, so students do their best.

The assessment also has built-in flexibility for pausing and resuming a test to accommodate interruptions or to allow students to complete a test over multiple shorter sittings if needed.

Due to the adaptive nature of the assessment, most students are expected to respond correctly to approximately 50 percent of the items administered (see Section 5.3.4 for more information).

3.3.2.2 Test Expiration Rules

To support valid assessment data interpretation and use, it is recommended that all students in a district, school, and grade complete the test within the same two- to four-week testing window. An assessment remains active in a student's account for a limited time before expiring. Administrators and proctors should monitor assessment status to ensure that students complete each assessment promptly once they start.

A test event expires 28 days after a student begins the assessment. If a test is due to expire and the student is actively working on it, the student can complete the test. However, if the student exits the test before completing it, the expiration will still occur at the 28-day mark. When a test expires, a new test event is automatically reassigned.

If a student has seven days or fewer remaining to complete an in-progress assessment, a red hourglass icon appears next to the student's name on the Inform Status (Reading Spanish) report and teacher dashboard, indicating the number of days remaining before expiration.

3.3.2.3 Monitoring Student Progress and Activity

During test administration, proctors are responsible for actively monitoring student progress and engagement. They observe students to ensure they remain on task and encourage them to put forth their best effort.

The *i-Ready* platform includes real-time tracking that allow proctors to monitor each student's test status and expiration date. If a student appears to be progressing through the test unusually quickly, teachers have the option to reset a test in progress.

3.3.3. After the Test

3.3.3.1 Posttest Procedures

Rush flags indicate when a student may have completed the assessment more quickly than expected, suggesting insufficient effort. These flags are based on test-level timing data and appear next to the Overall Placement & Scale Score in the *i-Ready Inform* Results Report (Class).

Rush flags will be available beginning in the 2026-2027 school year. Test administrations from the 2025-2026 school year will be used to collect test-level timing data to develop the yellow Rush flag. Based on this year's data, the top 5% of fastest finishers will be flagged as potentially rushed.

If a student completes the test in less than a specified amount of time, a yellow flag will appear next to the Overall Placement & Scale Score, signaling that a student may have rushed through the test. This alert prompts educators to evaluate the student's test results alongside other indicators of performance. Depending on district policy, the student may be asked to retake the test; however, the original results remain in the system and are not overwritten.

Unlike *i-Ready Inform*, Rush flags for *i-Ready Inform de lectura* are currently based on test-level data rather than item-level data. As more data become available, item-level analysis may be used, and red Rush flags could be introduced in future updates.

3.3.3.2 Reviewing Results, Setting Goals, and Basing Instructional Decisions on Results

After the test, results are reported to educators to support instructional decision making, as outlined in the Theory of Action in Chapter 1. Teachers are encouraged to review assessment results, alongside other performance information, with students to set academic goals as they gain reading proficiency. Score reports highlight areas of strength and recommend instructional priorities.

During the 2025–2026 school year, *i-Ready Inform de lectura* results will not be directly linked to an automated Personalized Instruction My Path for Spanish Reading. However, teachers can manually assign Spanish Reading lessons (Grades K-5) to support students' instructional needs.

3.3.4. Resources and Support Materials for Administrators and Proctors

A comprehensive library of resources and support materials is available on [Success Central](#). These resources support smooth and effective test administrations. The site offers frequently asked questions (FAQs) about the entire *i-Ready* product suite, brief training videos, helpful reference sheets, and many other resources. The document titled [Proctoring Guidance for the Inform](#) provides helpful recommendations about what to do before, during, and after test administration. The [Get Good Inform Data](#) page provides checklists and other tools to ensure reliable data about student performance are obtained from assessments. Testing Windows and Administration Time

3.3.5. Testing Windows

i-Ready Inform de lectura is typically administered three times per academic year: fall, winter, and spring, with additional administrations as needed. Table 3.1 shows the recommended dates for each testing window **Error! Reference source not found..**

While these windows provide a structured schedule, districts may set testing dates within these windows to meet their unique scheduling needs. Testing is available year-round, allowing flexibility in administration.

Table 3.1. <i>i-Ready Inform de lectura</i> Recommended Testing Windows			
School Year	Fall	Winter	Spring
2020–2021 and beyond	August 1 to November 15	November 16 to March 1	March 2 to June 15

3.3.6. Administration Time

Because the 2025-2026 school year is the first administration year for *i-Ready Inform de lectura*, data on testing times has not yet been collected. Therefore, recommended administration times are based on those for *i-Ready Inform*. These testing times and session lengths are planning suggestions; actual student needs may vary based on multiple factors.

Although the assessment is not timed, districts are advised to plan testing sessions using the recommended total testing times for each grade level, shown in **Error! Reference source not found.** Proctors should allow extra time for transitions, logging in, viewing tutorial videos, and supporting students who were absent or need more time to complete the assessment before it expires.

Curriculum Associates recommends delaying Grade K testing by four to six weeks to give students time to adjust to the classroom experience and become familiar with using technology. Resources are available to help educators prepare students accordingly.

Table 3.2. <i>i-Ready Inform de lectura</i> Recommended Schedule by Grade		
Grade	When to Schedule the First <i>i-Ready Inform de lectura</i>	Recommended Testing Session Times*
K	Four to six weeks into the school year	Three 20-minute sessions
1	Start as soon as possible	Two 20- to 30-minute sessions
2–6	Start as soon as possible	Two 40- to 50-minute sessions

Note: Recommended testing session times do not necessarily reflect actual testing times, rather they are provided to help support scheduling and administration.

3.4. Accessibility and Accommodations

To meet the needs of the students and districts we serve, we continuously improve the accessibility of our educational tools and resources. Consistent with our Accessibility and Inclusion Framework (see Chapter 2), we have a systematic approach to accessibility that includes:

- applying Web Content Accessibility Guidelines (WCAG) and the Universal Design for Learning (UDL) framework to guide our efforts
- engaging an internal team of experts in access and equity, curriculum, assessment, policy, and research expand usability for diverse learners
- conducting rigorous reviews with external accessibility experts to align with established and evolving best practices
- incorporating guidance and feedback from the districts and educators we serve

In addition, we enlist student testers to evaluate the effectiveness of accessibility enhancements. Our commitment to accessibility is ongoing, and all enhancements are included in Curriculum Associates' Software-as-a Service (SaaS) model.

i-Ready Inform de lectura includes multiple levels of accessibility features:

- **Universal supports** are available to all students, regardless of disability status or any other type of need and are designed to make the assessment broadly inclusive.
- **Designated supports** are available based on decisions made by educators or teams familiar with a student's characteristics and needs.
- **Accommodations** are provided in accordance with a student's individualized education plan (IEP), 504, or English Learner plan.

Some accessibility features are embedded directly within the *i-Ready* platform, while others are provided externally(non-embedded).

For complete information, educators can visit Curriculum Associates' publicly available [Accessibility and Accommodations Resource Hub](#). This hub includes the [i-Ready Inform Universal Accessibility Features and Accommodations Guidance Document](#), which offers comprehensive guidance for *i-Ready Inform de lectura*.

3.5. Test Security and Privacy

By implementing comprehensive monitoring procedures, prioritizing platform security, safeguarding data, and establishing effective policies and oversight, the test administration process preserves assessment integrity, protects data privacy, and maintains stakeholder trust and confidentiality of stakeholders.

3.5.1. Supervision

Ensuring the integrity and security of test administration is essential to maintaining the reliability of assessment results and validity of its intended interpretations and uses. To support this, Curriculum Associates provides training and detailed guidance on scheduling, student engagement, and monitoring throughout the testing process. School leaders, teachers, or other designated proctors are responsible for supervising the testing environment, actively monitoring test-takers to prevent irregularities, and reviewing status reports to ensure timely completion. Proctors also enforce adherence to test administration protocols and offer test-takers encouragement or breaks as needed, helping to maintain a fair and controlled testing environment.

3.5.2. Secure Browser

Curriculum Associates offers an optional secure browser solution that prevents students from printing, copying, accessing other websites, or using other applications while taking *i-Ready Inform de lectura*. The feature is designed for educators wanting increased security during the administration. Implementing secure testing will:

- lock the testing environment when students begin the assessment
- give educators confidence that students are on task by minimizing distractions
- ensure the accuracy and reliability of assessment data

The secure browser can be enabled for an entire district, or for specific schools. Once the secure browser is enabled, all students accessing *i-Ready* from the enabled district or school must use the secure browser when taking *i-Ready Inform de lectura*.

3.5.3. Data Transmission and Storage

i-Ready provides various layers of security. We conduct data transmission for *i-Ready* using a secure internet protocol (e.g., Secure Sockets Layer [SSL] and Hypertext Transfer Protocol Secure [https]). As a web-based, SaaS application, *i-Ready* is accessible only through https. In cases where we need to transfer data with a district student information system, we have alternative secure technologies available. *i-Ready* uses the following types of encryption:

- *Database encryption:* All *i-Ready* database servers are encrypted using industry standard AES-256 encryption.
- *Network encryption:* *i-Ready* is accessible only via https, and all public network traffic is encrypted with the latest encryption standards.
- *Data-at-rest/data-in-motion encryption:* Encryption of data at rest is implemented for all data stored in the *i-Ready* system. All data transfers are conducted using a secure internet protocol (e.g., SSL and https).

- *Backup encryption:* As indicated above, all *i-Ready* database servers are encrypted using industry-standard AES-256 encryption.

Curriculum Associates hosts all our servers with Amazon® Web Services.

Amazon® Web Services and all related marks, including logos, graphic designs, and service names, are trademarks or trade dress of AWS in the US and other countries.

3.5.4. Data Security and Privacy

Our Data Handling and Privacy Statement—posted at [CurriculumAssociates.com/Support/Privacy-and-Policies/i-Ready-Data-Handling-Privacy](https://www.curriculumassociates.com/support/Privacy-and-Policies/i-Ready-Data-Handling-Privacy) <https://www.curriculumassociates.com/support/Privacy-and-Policies/i-Ready-Data-Handling-Privacy>—describes our current data security policies and practices. It explains how we safeguard the sensitive information in *i-Ready*.

Chapter 4. Computerized Adaptive Testing

4.1. Chapter Summary

i-Ready Inform™ de lectura is a computerized adaptive test (CAT) with an item selection algorithm that maximizes measurement precision while ensuring that students are assessed on the appropriate content. This chapter includes an overview of the core components of the CAT: the vertical scale, content balancing across domains within each grade, item selection, stopping rules, methods for estimating student proficiency, and the item bank. The chapter concludes with a summary of simulation research Curriculum Associates conducts to refine and improve processes and the effectiveness of *i-Ready Inform* de lectura.

4.2. Introduction

Computerized adaptive testing is an assessment administration method defined by its ability to tailor the items administered to each student based on their responses to previous items. *i-Ready Inform* de lectura leverages CAT in conjunction with grade-specific and, in some cases, proficiency-specific test flows, which specify the number of items administered from each domain (described later in this Chapter). The assessment provides a strong degree of measurement precision across the entire proficiency spectrum, ensuring that instructional resources recommended based on results are appropriately aligned with students' needs.

4.2.1. Vertical Scale

i-Ready Inform de lectura reports student proficiency on a vertical scale. This scale is separate from *i-Ready™ Inform* for Reading and *i-Ready™ Inform* for Mathematics scales. Regardless of a student's chronological grade level and the items administered, scores are reported on the same 200–500 score scale. Critical components of the vertical scale—field testing, item calibration methodology, scale transformation constants, item drift monitoring, and scale maintenance—are discussed in Chapter 5.

4.2.2. Computerized Adaptive Testing

CAT programs must achieve a balance between item selection methods that support measurement precision while meeting specific content requirements and managing item exposure. *i-Ready Inform* de lectura balances content representation and measurement efficiency. This balance is achieved by structuring the test to present questions from all content domains without requiring the administration of items that are likely incongruent with the student's proficiency level, minimizing testing time. By taking advantage of CAT, *i-Ready Inform* de lectura administers items that closely

aligned with students' proficiency levels, resulting in a shorter test than would be possible with a fixed-form (i.e., non-adaptive) assessment.

i-Ready Inform de lectura begins with a starting proficiency estimate. The methods for selecting the starting items are discussed later in this chapter. After the administration of the first item or testlet, the item selection algorithm compiles a pool of eligible items for that student and selects an item appropriate for that proficiency level. As the student progresses through the assessment, the system adapts in real time, selecting items based on the student's performance up to that point. The CAT continues to progressively narrow in on the student's proficiency level, until the stopping rule is satisfied.

Ideally, the items selected for students during an adaptive assessment should be neither too easy nor too difficult, but rather appropriately challenging for that student to maximize measurement precision. In general, students who answer items correctly receive more difficult items, while students who answer items incorrectly receive less difficult items until the algorithm identifies each student's proficiency level.

The five major components of our CAT are:

- content balancing as accomplished through test flows
- adaptive item selection logic
- proficiency estimation
- operational item bank
- test termination criteria

Each of these is discussed in detail in the following sections.

4.3. Test Flows

Content balance and representation is established by the pre-defined *test flow* for each grade. The test flow specifies the order in which domains are tested and the number of items administered for each domain. The domains assessed may vary by grade level and interim proficiency estimates during the assessment. Results for each assessment completed at the grade do not impact the test flow for tests administered for the remainder of the grade.

4.3.1. Reading Test Flows

The design of *i-Ready Inform de lectura* supports accurate assessment of a student's Spanish reading proficiency. All students, regardless of grade level, are presented with items from the Vocabulary and Comprehension domains. Students in lower chronological grade levels, or performing at lower proficiency levels, may also be assigned items from the Phonics and Phonological Awareness domains. Reading test flows route students into these foundational skill domains if the interim proficiency

estimate determines that these additional domains are appropriate to measure. The test flows are shown in **Error! Reference source not found.** 4.1, Figure 4.2, and Figure 4.3..

Figure 4.1 illustrates the Grades K-1 test flow for Spanish reading. Students are administered Phonics items first, then Phonological Awareness items, followed by Vocabulary items, and lastly Comprehension testlet-based items.

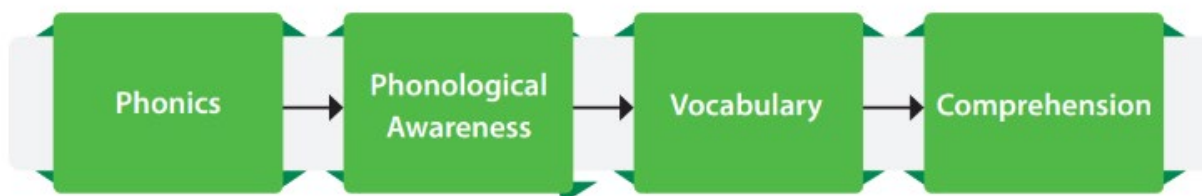


Figure 4.1. Reading Test Flow—Grades K and 1

Figure 4.2 illustrates the Grade 2 test flow for Spanish reading. Students receive Phonics items first, then Phonological Awareness items if the student has an estimated proficiency less than 303 (and would benefit from being assessed in that domain). All students then receive items from the Vocabulary and Comprehension domains.

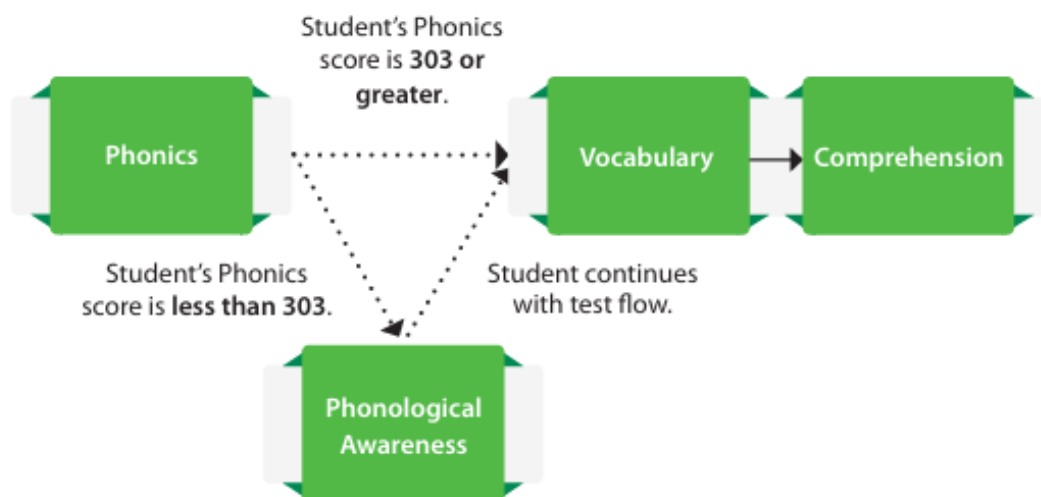


Figure 4.2. Reading Test Flow—Grade 2

Figure 4.3 presents the test flow for Spanish reading for students in grades 3–6. Students receive Vocabulary items first, then Comprehension items. Students receive Phonics items only if their overall score is below 382.

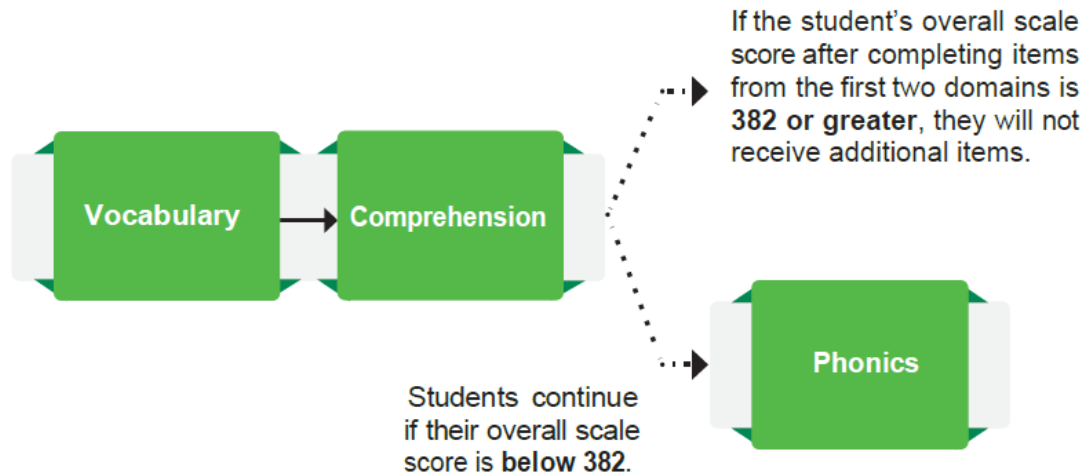


Figure 4.3. Reading Test Flow—Grades 3–6

Table 4.1 summarizes the number of items in the test flow by domain and grade level.

Table 4.1. Reading Items by Domain, Grade, and Test Flow					
Domain	Grades K–1	Grade 2		Grades 3–6	
	—	Phonics Score < 303	Phonics Score ≥ 303	Overall Score < 382	Overall Score ≥ 382
Phonics	12	12	12	10-12	—
Phonological Awareness	12	12	—	—	—
Vocabulary	10-12	10-12	10-12	14-16	14-16
Comprehension	20-24	20-24	20-24	28-32	28-32
Total Operational Items	54-60	54-60	42-48	52-60	42-48

4.3.2. Grade-Level Content Restrictions

4.3.2.1 Grade-Level Caps

The vertically scaled item bank allows for the administration of on-grade level items, which match the student’s chronological grade, and certain off-grade level items when needed to measure the student’s proficiency accurately. Off-grade level items support measurement precision across the proficiency spectrum, but students must also receive items that assess appropriate content. Therefore, we set a maximum allowable item grade level for each chronological grade and domain, referred to as a **grade-level cap**.

For Spanish Reading, the grade-level cap is set at three grade levels above the student's chronological grade. For example, a Grade 2 student administered a Spanish reading assessment could be served items from the bank representing Grades K–5. The grade-level cap is set at three grade levels above, so students in Grades 3–6 will all have the same available item bank.

4.4. Adaptive Item Selection

4.4.1. Starting Items

Selection of the first item depends on whether the student has previously taken *i-Ready Inform de lectura*. For students without a prior score, the system randomly selects an item from a predetermined set of items with a difficulty level closest to the designated starting location for that grade, which is one placement level below the student's chronological grade. These items are also eligible to be administered during the remainder of the assessment if not used as the starting item. For students who have previously taken *i-Ready Inform de lectura*, the system selects the starting item using the standard item selection rules outlined in the next section.

4.4.2. Item and Testlet Selection

Under the Rasch (1960) model, measurement information (i.e., test score reliability) is maximized when item difficulty is equal to student proficiency. As item difficulty and student proficiency diverge, information for that student decreases. Consequently, items that are too easy or too difficult for a student provide little information about the student's proficiency level. Rather, selecting an item that minimizes the difference between item difficulty and student proficiency, a method known as *b-matching*, optimizes precision of proficiency estimation at that point in the test. The remainder of this section uses the term *items* to refer either to a single item or to a testlet that contains a set of items.

i-Ready Inform de lectura relies on a Relaxed CAT item selection method (Cappaert, Morrison, and Su, 2022). This method balances high quality measurement properties with the flexibility needed to select items that give students more opportunities to demonstrate proficiency in the content they are learning in the classroom, while also controlling exposure to repeated items. The Relaxed CAT begins with a b-matching procedure that identifies items with the smallest difference between the student's estimated proficiency and each item's difficulty. The difference between student proficiency and item difficulty represents an undesirable attribute: the larger the difference, the less desirable the item is for administration.

The eligible item bank at any point in the test is based on:

- the student's grade level
- the domain from which the next item will be selected, as defined by the test flow
- item grade-level caps

To minimize the number of repeated items across administrations for a student, a set of penalty weights is applied to items that were previously administered to the student at any point in the past 365 calendar days. Items and testlets are never repeated within an administration but could be repeated across test administrations in limited situations. Penalty weights reduce—and often eliminate—the likelihood of repeated items being selected and served to students during the assessment. They work by applying a weight to increase the difference between item difficulty and student proficiency, which makes the item appear less informative and therefore less likely to be chosen. The system also increases penalty weights for items administered on recent tests, when item recency effect would be strongest. These penalties have undergone rigorous simulation testing and are applied at both the grade and domain levels.

To give students more opportunities to demonstrate proficiency on grade-level skills taught in the classroom, the system applies an additional set of penalty weights based on the difference between the item’s grade level and the student’s chronological grade. These penalties may differ by domain. When an item is above or below a student’s grade level, the penalty weight increases the difference between the item’s difficulty and the student’s proficiency. This makes the item appear less informative and decreases its likelihood of selection. The penalty weight increases as the discrepancy between item grade and student grade increases. As the penalty weight increases, the likelihood of an item being selected decreases, assuming everything else is held constant. Items that match the student’s chronological grade are not penalized, and items near the student’s grade level receive low (or no) penalty weights to increase their likelihood of being selected. Grade-by-domain penalties for above-grade content limit exposure to higher-level content unless a student’s performance indicates that exposure is appropriate.

The Relaxed CAT method produces a single summed penalty weight for each eligible item in the item bank each time an item is selected. The item with the smallest total penalty is selected. This total penalty is calculated by summing the absolute difference between student proficiency and item difficulty from the b-matching step, the repeat-item penalty weight, and grade-level penalty weight. If multiple items share the smallest total penalty weight, the system randomly selects one of the items to administer.

In most Spanish reading domains, items are selected one at a time using this procedure. For the Spanish reading Comprehension domain, the system uses the same process to select whole testlets based on the testlet’s difficulty, defined as the average difficulty of the items in the testlet.

4.4.3. Domain Termination Logic

Domain termination rules differ based on whether the domain has individual items or testlets. Table 4.1 references the range of items (minimum and maximum) from each domain for each grade level.

For domains with individual items—Phonics, Phonological Awareness, and Vocabulary—domain termination occurs when 1) the maximum number of items for the domain has been reached, or 2) when the student has been administered the minimum number of items for the domain, as defined by

the test flow, and the student's standard error of measurement (SEM) for that domain is below a predetermined threshold.

For the Comprehension domain, which uses testlets, the domain terminates under one of two conditions: 1) the maximum number of items for the domain has been reached, or 2) when the student has completed the minimum number of items for the domain, as defined by the test flow and has finished all items in the testlet, which would require selecting another testlet to continue.

This logic ensures that students receive enough items from each testlet to support high-quality measurement. It also prevents situations in which a student reads a lengthy passage but answers only one or two items from the associated testlet. In this way, the process balances measurement precision with efficient use of student testing time.

4.4.4. Test Termination Logic

The test terminates when a student has completed each domain according to the relevant termination criteria described in the prior section. **Error! Reference source not found.** shows the possible operational item counts—depending on the test flow—by grade and subject. The total test length could be slightly longer if field-test items are included (see Chapter 5 for details).

4.5. Proficiency Estimation

4.5.1. Starting Proficiency

The starting student proficiency depends on whether the student has a prior *i-Ready Inform de lectura* score stored in the assessment system.

For students who have not previously taken *i-Ready Inform de lectura*, their starting proficiency estimate is set to the overall assessment cut score for the placement level one grade level below their chronological grade. For example, a Grade 4 student would have their starting proficiency estimate set to a scale score of 382, which is the lower bound of a Grade 3 placement level for a Grade 4 student.

For students who have previously taken *i-Ready Inform de lectura*, their starting proficiency is their overall proficiency estimate from their most recent assessment.

4.5.2. Proficiency Estimation

4.5.2.1 Overall Estimates

Student proficiency estimates, based on the Rasch (1960) model, are updated after each item response using a Maximum-Likelihood IRT estimation approach. These estimates are used for subsequent item selection. Maximum-Likelihood identifies the proficiency value that maximizes the probability of the observed response pattern (correct/incorrect responses) given the items

administered to that student on that test (and the associated item difficulties). An iterative procedure refines the proficiency estimate to maximize this probability. When the change from one estimate to the next is sufficiently small, the estimate is considered converged and further iterating to improve the estimate is not required.

Estimates are updated both within the test (interim proficiency estimation) and at the end of the test (final proficiency estimation). To control the magnitude of adjustments to the proficiency estimate, particularly near the beginning of an assessment, the maximum allowable adjustment after the administration of any single item is limited to ± 1.0 logits. Generally, the magnitude of adjustment to a student’s proficiency estimate decreases as the student progresses through the test as the estimation algorithm progressively narrows in on the student’s proficiency.

4.5.2.2 Domain Estimates

The procedure for domain-level proficiency estimation is the same as for estimating the overall score, except that the estimate is calculated using only the items within the domain defined in the test flow for that grade. The student’s starting domain-level proficiency estimate is set equal to the student’s overall proficiency estimate. Domain-level estimates are then updated after each item and again at the end of the domain. Domain-level estimates can be used for routing decisions in certain test flows, following the same estimation process used for the overall score.

4.5.2.3 Extreme Score Adjustment

In the case of extreme score patterns—either all correct or all incorrect at the overall or domain level—a slight adjustment is made to the scoring algorithm to produce a finite estimate.

4.6. Item Bank

As of fall 2025, the *i-Ready Inform de lectura* item bank has 1,571 Spanish reading items. The operational item bank will continue to grow as new items are field tested and evaluated for inclusion in the bank.

Table 4.2 displays the number of items assigned to each grade and eligible for administration based on the grade-level caps described earlier. Table 4.3 shows item counts by domain.

Table 4.2. Number of Items by Grade		
Grade	Spanish Reading	
	By Grade	Eligible
K	294	992
1	264	1210
2	234	1414
3	200	1571

4	218	1571
5	204	1571
6	157	1571

Table 4.3. Number of Items by Domain			
Phonological Awareness	Phonics	Vocabulary	Comprehension
121	139	444	867

Test information functions (TIFs) are unique to each student and are based on the items administered; therefore, they cannot be easily presented here. However, information functions that are averaged across all eligible items for a grade level or grade band can be used to understand the nature of the item bank. **Error! Reference source not found.** shows the average item information curve at each grade level. These curves show the degree of information across a broad range of proficiencies. Additional information and results, including details on student-level measurement precision, are presented in Chapter 6.

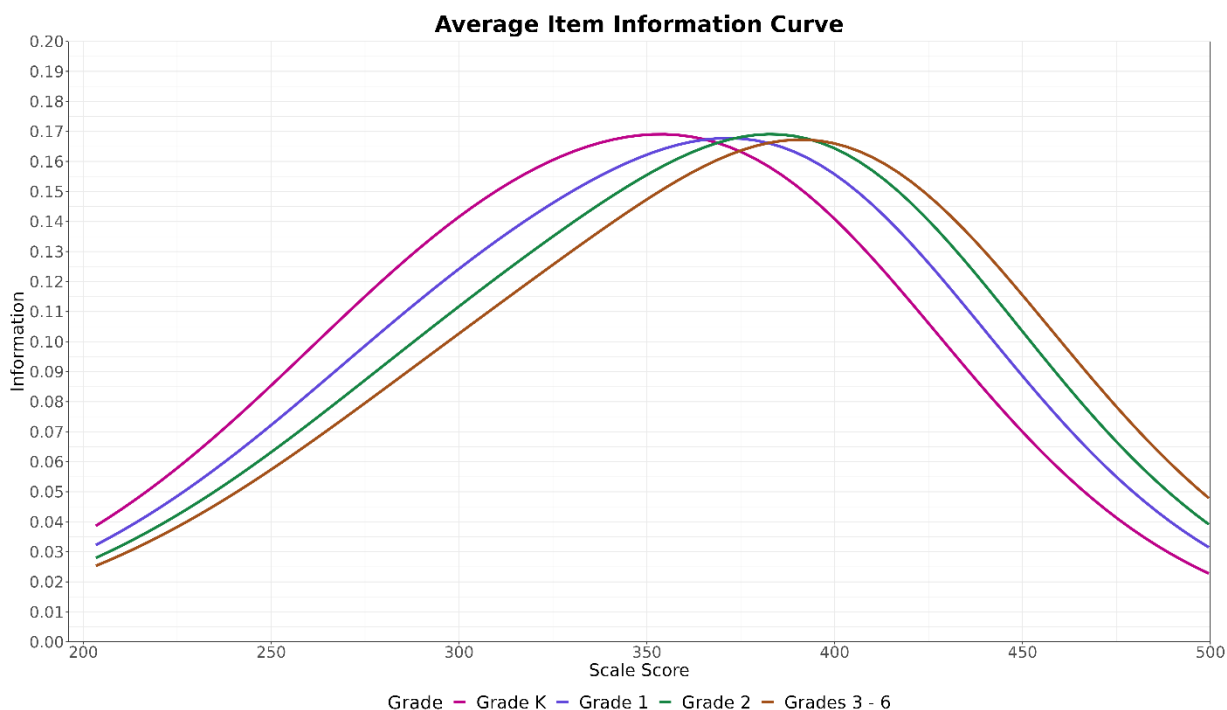


Figure 4.4. Average Item Information Curve by Grade for Spanish Reading

The distribution of items in the bank for *i-Ready Inform de lectura* is appropriate for an adaptive assessment. Figure 4.5 presents the distribution of all items as a histogram in relation to an estimate of

the Grades K–6 student population density based on student data collected between August and mid-October of the fall 2025 testing window. This figure demonstrates that item and student proficiency distributions are mostly aligned. The figure also shows that item development has produced a substantial number of difficult items, allowing for reliable measurement of high-proficiency students and enough items to measure student proficiency across fall, winter, and spring testing windows. Item development will continue to increase the number and variety of items in the *i-Ready Inform de lectura* item bank. This figure will be generated again after testing is completed for the 2025-2026 administration year to show the student proficiency distribution for the full year.

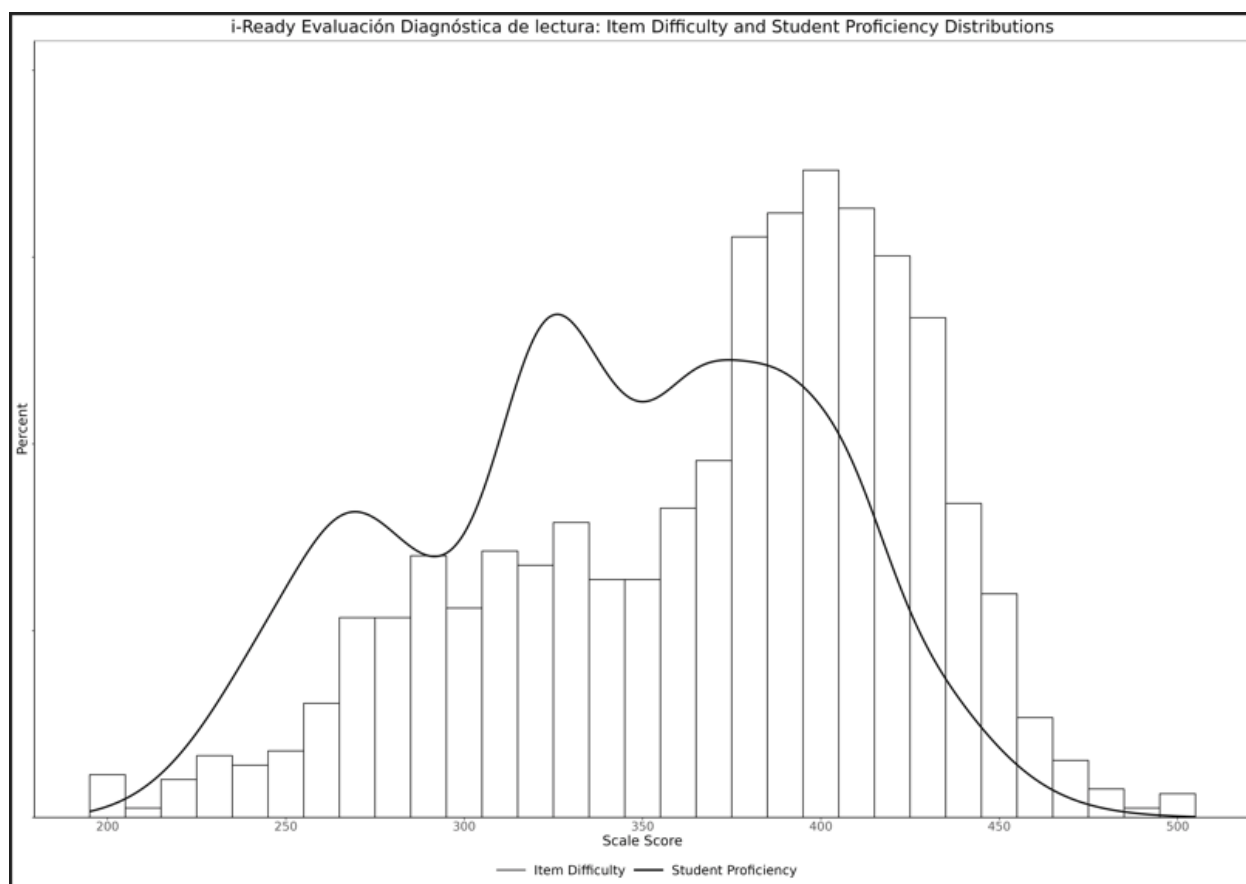


Figure 4.5. Item Difficulty and Student Proficiency Distributions for Spanish Reading

4.7. CAT Simulation Studies

Curriculum Associates uses simulations to refine its processes and improve the effectiveness of its assessments. A custom-developed simulator is used to evaluate the efficiency of decisions regarding item selection, proficiency estimation, and item development priorities.

4.7.1. Simulator Structure and Capabilities

The CAT simulator is a modular tool, mirroring the structure of the operational CAT algorithm to allow for accurate simulations. It features interchangeable modules such as the item bank, student proficiency distribution, student proficiency estimation method, test flow, and item selection methods. This approach allows for simulating changes being considered, while maintaining the integrity and functionality of the CAT algorithm.

The core advantage of this simulator is its capacity to evaluate changes to the item bank, item selection methods, and test flows, as well as other variables. It measures and provides detailed analyses of mean conditional standard error of measurement (CSEM), root mean squared error (RMSE), and bias. (See Chapter 6 for the definition of CSEM and the associated results.)

4.7.2. Simulator Outputs and Usage

Simulation study outcomes play a central role in maintaining the *i-Ready Inform de lectura* item bank. Simulations are conducted using simulated data for representative normal proficiency distributions and uniform proficiency distributions to reflect typical students and those in the tails of the proficiency distribution. These studies help Curriculum Associates understand how changes to key parameters may affect student assessment outcomes. Simulations serve as a validation tool for refining the item bank, adjusting test flows, improving item selection methods, and informing item development. Through these data-driven insights, Curriculum Associates continually enhances its assessments to provide more precise and meaningful results for all students.

Chapter 5. Calibration and Scaling

5.1. Chapter Summary

Chapter 5 presents the psychometric theory and supporting research on which *i-Ready Inform*™ de lectura is built. This chapter presents evidence for the use of key validation to evaluate items for inclusion in the item bank, calibration procedures and development of the vertical scale for the initial item bank, future field test plans, and scale score transformation equations. These procedures demonstrate our use of best-practice approaches to build the score scale to establish a reliable assessment for measuring and reporting student proficiency in the Spanish reading content domains assessed by *i-Ready Inform* de lectura.

5.2. Introduction

Curriculum Associates uses industry-standard methods rooted in both classical test theory (CTT) and item response theory (IRT) to develop and maintain the *i-Ready Inform* de lectura item bank. This chapter describes the procedures used to establish the item bank through field testing new items, estimating item parameters, and developing a vertical scale. The transformation of scores from the underlying vertical scale to the reporting scale is also described.

5.3. Key Validation

Before appearing on an operational test form, all items undergo field testing to ensure they meet high quality standards. During the key validation process, psychometricians and subject matter experts review key item statistics to identify any issues that may require further evaluation. This validation step confirms whether each item performs as intended and is suitable for operational use.

5.3.1. Classical Test Theory

Classical test theory (CTT) provides a statistical model for the number of items students answer correctly on a test. When model assumptions are satisfied, CTT facilitates the evaluation of test score precision (e.g., reliability and the standard error of measurement). In CTT, student performance is generally represented by the total number of items answered correctly.

CTT analyses assume that an observed score on a test can be modeled as a sum of two components: the student's "true" (or expected) score and residual error specific to the observed score. In other words, CTT assumes that students possess stable characteristics that manifest in observed test scores, which can be represented mathematically as

$$X_i = T_i + e_i \quad (1)$$

where X_i is the observed score for respondent i , T_i is the respondent's true score on the test, and e_i is random error that influences the expression of respondent i 's true score.

5.3.2. CTT Indices used for Evaluating Items

Key validation uses a variety of indices rooted in CTT. Extreme or unexpected values can help identify items that deserve closer scrutiny. However, any items identified based on these criteria are not automatically rejected or retained.

5.3.2.1 Item Difficulty

The p-value is a sample-dependent measure of item difficulty and is calculated by dividing the number of students who responded correctly to an item by the total number of students who attempted the item. P-values range from 0.0 to 1.0 with higher values representing a greater proportion of students answering the item correctly (e.g., the item is easier). They help identify items that may be too easy or too difficult for a given group of students, identifying potential flaws or clues of the correct answer choice in the item. Evaluating p-values helps ensure the item bank includes high quality items that span the full range of item difficulties, which is necessary to precisely measure the proficiency of a wide range of students.

5.3.2.2 Item distractor response proportion

The proportion of students who select each distractor—an incorrect answer choice—provides insight into how students interact with an item. Distractor response proportion is computed for each distractor and is calculated as the number of students who selected the distractor divided by the total number of students who answered the item. If this statistic is computed for the keyed correct response, the p-value for the item is obtained, as described in the previous section. Ideally, a higher proportion of students select the key (correct answer) compared to distractors so that distractors have lower response proportions. Distractors with response proportions greater than the response key are scrutinized.

5.3.2.3 Item discrimination

The item-total correlation (i.e., the point-biserial correlation) is the correlation between the item score (incorrect/correct) and the total test score. This measure, which ranges from -1.0 to 1.0, indicates how well an item discriminates between low- and high-performing students. Positive values indicate that students with higher scores on the overall test are more likely to select the response key (expected), while negative values indicate that students with higher scores on the overall test are less likely to select the response key (not expected). Values near 0 indicate little to no relationship between a student's score and selecting the response key. Higher-performing students should be more likely to select the key than lower-performing students, resulting in a positive correlation. A range of positive point-biserial correlations is desirable, although higher values are generally preferred because items with high, positive discrimination better differentiate among students with different scores.

5.3.2.4 Item distractor-total test correlation

The item distractor-total test correlation, which is also a point-biserial correlation, is the correlation between each distractor and the total test score and ranges from -1.0 to 1.0. This is referred to as item discrimination when computed on the response key. This value is also computed for each distractor.

Positive values indicate that students with higher scores on the overall test are more likely to select the distractor (not expected), while negative values indicate students with higher scores on the overall test are less likely to select the distractor being evaluated (expected). Values near zero indicate little to no relationship between a student's score and selecting the distractor being evaluated. A negative point-biserial correlation is expected for distractors, indicating that the higher the student's score, the less likely the student is to have chosen the incorrect response.

Items with positive distractor-total correlations are scrutinized. However, a weak positive correlation for a distractor does not necessarily indicate that the item is problematic, especially if the distractor correlation is substantially smaller in magnitude than the item-total correlation. In some cases, it may simply mean that the distractor is attractive to moderately-performing students and unattractive to low-performing students.

5.3.3. Item Evaluation

Data are used to compute classical item statistics to evaluate if items are functioning consistently with expectations. The data used to calculate these statistics depends on the purpose of the item evaluations. For example, field test data may be preferred when a wide representation of student proficiency is desired. Items with potential psychometric issues are flagged for review. For example, items may be flagged for very low or very high p-values, near zero or negative discrimination values, distractors selected by a high proportion of students, or a high proportion of high-achieving students selecting a distractor.

Items flagged by the psychometrics team are thoroughly reviewed by subject matter experts (SMEs). To aid in this evaluation, SMEs are provided with IRT difficulty parameter estimates for the items. SMEs verify the accuracy of the item and determine if the item is functioning as expected and should be accepted for the item pool, if the item should be revised and re-field tested, or if the item should be removed. For example, adaptive assessments require items of extremely low difficulty and extremely high difficulty to adequately measure students of low and high proficiency. In this case, the SME review might result in an item with extremely low or extremely high difficulty being retained for the item pool because the item is functioning as expected.

5.3.4. Rasch Model

Curriculum Associates follows industry-standard methods for calibrating and maintaining score scales for *i-Ready Inform* de lectura. Items are calibrated using the Rasch (1960) measurement model. A unidimensional Rasch model assumes that the assessment measures one underlying construct—

Spanish reading in the case of *i-Ready Inform de lectura*. In the Rasch model, the probability of a correct response for student i on item j , P_{ij} , is modeled as

$$P_{ij} = \frac{e^{(B_i - D_j)}}{1 + e^{(B_i - D_j)}} \quad (2)$$

where B_i is the proficiency estimate of the student, and D_j is the difficulty parameter of item j . A key advantage of this model is that item difficulty and student proficiency estimates are placed on the same scale.

The difficulty parameter D_j in the Rasch measurement model is the location on the scale at which a student with a proficiency estimate equal to the item difficulty has a 0.5 probability of answering the item correctly. The difficulty parameter also corresponds to the inflection point of the item response function where that function has the steepest slope and where the item provides the most information about student proficiency. This principle guides the selection of items in a computer adaptive assessment like *i-Ready Inform de lectura*. For example, an item with difficulty (D_j) of 1.0 maximizes measurement information in the Rasch framework for a student with a proficiency B_i of 1.0. If such a student were administered a large set of items, all with item difficulty equal to 1.0, the student would be expected to answer half the items correctly, and the resulting proficiency estimate ($\hat{B}_i$) would be highly reliable.

5.4. Initial Item Bank Creation

Launching *i-Ready Inform de lectura* required the development of a vertically scaled item bank with items across all grades on the same scale. The development process leveraged the existing Assessment of Spanish Reading (ASR) item bank, comprised of approximately 1,000 items across Grades K through 6. After the vertical scale was created, additional items were field-tested, brought onto the vertical scale, and added to the bank to support the adaptive assessment.

5.4.1. Vertical Scaling Study

A vertical scaling study was conducted to associate performance at various test levels to a single score scale (Kolen & Brennan, 2010) and link items across student grade levels. The results of the vertical scaling study were used to bring all Spanish reading items onto the vertical scale.

5.4.1.1 Scaling Design and Forms

A **common item non-equivalent groups mixed design** was used for the vertical scaling study. The study forms for each student grade level contained common items, or linking items, to investigate student performance across grade levels (Briggs & Dadey, 2015). Each form was comprised of both on- and off-grade grade level items. Items were chosen such that each form satisfied a set of content and

psychometric criteria. For example, items were included on forms such that all domains available at that student grade level were represented. Criteria used to build forms and to ensure a set of diverse items included:

- At least 30% of the total test length for a form should be on-grade linking items; these items should be linked to other grade levels.
- The point biserial for an item should be greater than or equal to 0.2.
- The range of difficulty values for the lower-, on-, and upper-grade level items on a given form should have a heterogeneous distribution.
- A diverse set of item types (multiple-choice and technology-enhanced) should be included whenever possible.
- Content representation should be balanced.
- Items should not be clues to each other.

5.4.1.2 Calibrations

IRT places item difficulty and student proficiency metrics on the same scale. Each grade-level form was calibrated separately using the Rasch model and person-centering to put the items and persons on the same grade-level scale (Linacre, 2023). The base grade is considered the anchor point of the vertical scale. Grade 3 was chosen because it had a range of item logits, overlap between the item and person distributions, and would result in vertically scaled item difficulties that were not skewed in a single direction. For a final calibrated set of items, both item-level and person-level investigations were conducted. Items were flagged for review based on the key validation procedures above; these items were reviewed for potential removal from the common item set for a grade pair, of which few were removed. Examinees were also investigated for misfit and were subsequently removed from the analysis.

5.4.1.3 Scaling Constants

To calculate vertical scaling constants, the common item sets between adjacent grade pairs were used after dropping items and examinees as described earlier. The mean/mean method for creating constants was used between adjacent grade levels. In this method, the constant was represented by the difference between the means of the common item Rasch difficulties for the adjacent grades. These adjacent grade constants were used to create the final cumulative linking constants. The final constants from these calculations are presented in Table 5.1 and graphically in Figure 5.1.

Table 5.1. Final Vertical Scaling Constants	
Grade	Constant
0	-2.6559
1	-1.3794

2	-0.5420
3	0
4	0.2544
5	0.6237
6	1.0050

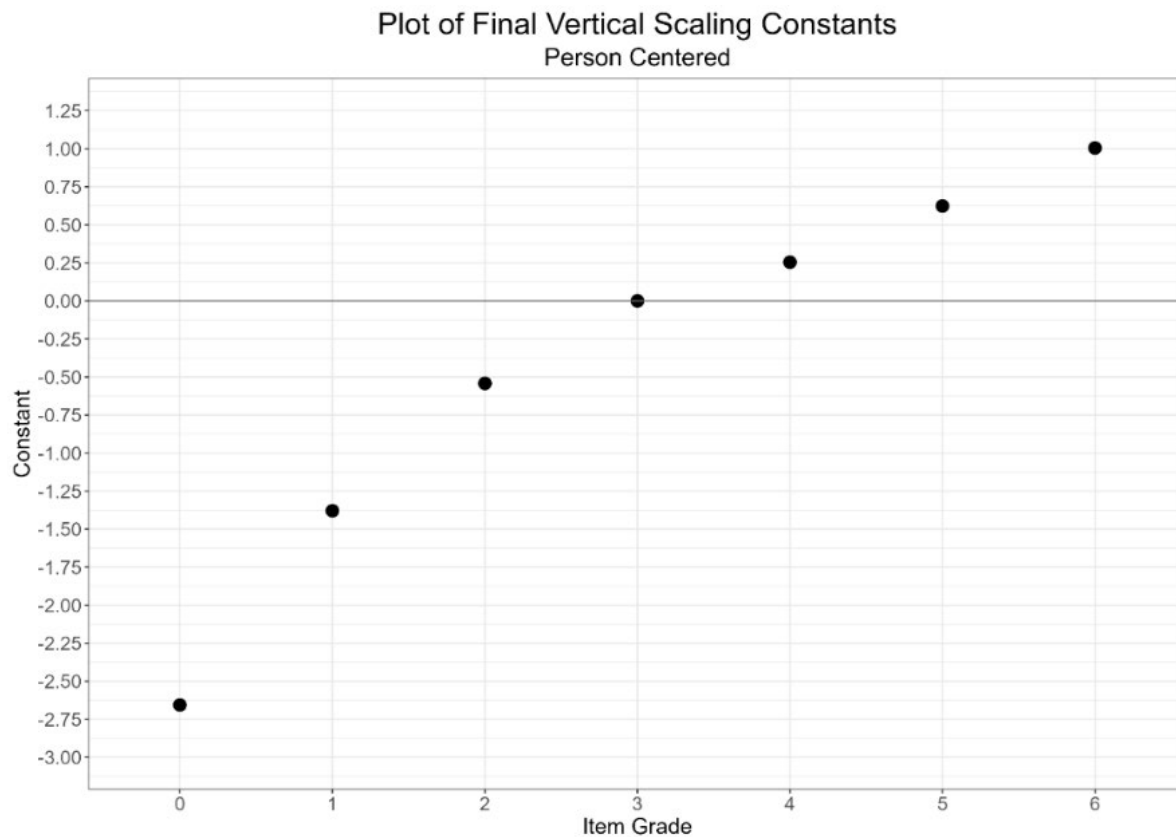


Figure 5.1. Final cumulative vertical scaling constants.

5.4.1.4 Evaluation

Several criteria were used to evaluate the vertical scale. Most of these criteria have been used in evaluations of other vertical scales. These criteria include:

- progression in difficulty across grades
- vertical scale means and standard deviations
- relationship between vertical scale item sets

All criteria were satisfied by the vertical scaling solution.

5.4.2. Standalone Field Testing

Field testing items is a necessary step in the test development process that allows data to be collected on newly developed items before they are used in operational administrations. To increase the size and robustness of the *i-Ready Inform de lectura* item bank prior to launch, standalone field test studies were conducted. Standalone field testing involves the administration of newly developed test items separately from an operational (or scored) administration of the test. This section describes the general IRT procedures used to field test items and transform those items onto the vertical scale.

5.4.2.1 Field Test Forms and Linking Design

Items were field-tested via standalone, fixed-form tests to on-grade level students. The field test forms contained items from all domains appropriate for that grade level (test flows are presented in Chapter 4). The field test design included common items, which allowed for a common Rasch scale to be established for all items across forms within grade level and later to be linked to the single vertical scale for Spanish reading.

5.4.2.2 Calibration

An IRT-based scale underpins the *i-Ready Inform de lectura* item bank and requires that new items be transformed to this scale when added to the item bank. A major advantage of this type of scale is that CAT programs can construct and deliver a test form by including a set of items previously calibrated to the bank scale. This design permits administering unique combinations of calibrated items to create custom test forms that optimize measurement precision for individual students.

Calibrations for the standalone field-test items involved fitting the Rasch measurement model to the item-level data using WINSTEPS (Linacre, 2023). Maximum likelihood estimation (MLE) was used to estimate Rasch parameters for all items and examinees in the sample. These calibrations set the scale such that each item grade level had its own horizontal scale. For example, all Grade 3 items were placed on the same scale but were not placed on the same scale as the Grade 4 items. The calibrations involved several iterations, and the same process was used across both testing windows. In the first iteration, responses from all students to all items were calibrated. Using this information, items underwent a key validation check. Items identified by SMEs were removed from the remaining analyses. Additional iterations were conducted to remove students with responses that did not fit the model and to obtain the final difficulty estimates for the items. After obtaining the final difficulty estimates for the standalone field-test items were brought onto the vertical scale as operational items using the same two-step process as described above for the remaining operational items.

5.5. Embedded Field Testing

Field testing will continue to be a key component in the test development process as new items are added to the *i-Ready Inform de lectura* item bank. After launch, the field-testing design will evolve from a standalone process to an embedded field test design. Embedded field test (EFT) items will be inserted into the test flow for eligible students. Students will only be administered these items when

new items are being developed, or currently operational items require evaluation; it is possible that a student's test will include no field-test items. These EFT items will not be used to calculate a student's score.

EFT items may be added to a student's operational test flow in two sections. First, students may receive a section of EFT items from the Phonics, Phonological Awareness, and/or Vocabulary domains. Students may receive a second section of items from the Comprehension domain. Items associated with a passage, or paired passages, are field tested together as a set.

A within-test check is conducted to determine if a student is eligible for administration of EFT items. Students performing well below grade level are not included in EFT data collection. For eligible students, the system compares the student's chronological grade to the assigned grades of the EFT items to determine which items are eligible for administration. An eligible student could receive EFT items that are on-grade level—meaning the item's assigned grade level matches the student's chronological grade level (e.g., a Grade 4 student receiving a Grade 4 EFT item), and/or also receive EFT items that are off-grade level, wherein an item could be one grade level below (e.g., a Grade 4 student receiving a Grade 3 EFT item) or one grade level above the student's chronological grade (e.g., a Grade 4 student receiving a Grade 5 EFT item).

Once a student is determined to be eligible to receive EFT items and the system has determined the EFT items eligible to be administered, an item will be randomly selected from the pool of eligible EFT items associated with a given section. Adaptive logic is not used; there is no attempt to match EFT item difficulty to the student's proficiency level. The number of EFT items is limited so any given student is not overly impacted. When items are based on passages, items associated with a passage are selected as a testlet and field tested together. Although EFT items may be administered to eligible students, the number of EFT items included in any given student's test will be based on both test development considerations and the impact to students.

5.6. Transformation to the Reporting Scale Score

i-Ready Inform de lectura uses linear transformations of proficiency estimates on the underlying logit scale for the vertical scale. The primary advantages of a linear transformation are simplicity and supporting score interpretation. A single transformation is applied across the entire scale. In addition, the distributions of the reported scale scores remain proportionally congruent with the distributions of the estimated $\hat{B}_i$ (or logit) scores, and the conditional standard errors are proportionally scaled relative to the slope constant and maintain the same convex, U-shaped pattern on both the $\hat{B}_i$ and reported score scale.

All estimates at both the overall test level and the domain level are transformed from the operational $\hat{B}_i$ scale to a scale score using the following linear transformation function:

$$\text{Unrounded Scale Score} = 396.2423 + (34.4475 * \hat{B}_i) \quad (3)$$

After student scores are scaled using the above transformations, they are rounded to the nearest integer. Floor and ceiling values are applied based on student grade level. Any scale score below 200 is replaced with the lowest observable scale score (LOSS) of 200 and any scale score above the values represented in Table 5.2 is replaced with the highest observable scale score (HOSS) for the student grade level.

Table 5.2. HOSS Values by Grade Level	
Student Grade	HOSS
K	458
1	473
2	473
3	500
4	500
5	500
6	500

Chapter 6. Reliability

6.1. Chapter Summary

This chapter reports marginal reliability estimates for the overall assessment as evidence of internal consistency of *i-Ready Inform™* de lectura. Estimates are reported for the total population and selected student groups. Reliability evidence reflects the cohesiveness and integrity of the test's internal structure, supporting that the test measures its intended constructs consistently with precision. The chapter concludes with evidence for classification accuracy and consistency.

6.2. Introduction

In educational measurement, **reliability** refers to the measurement precision of scores from an assessment. Reliability analyses provide an estimate of a test's internal consistency or measurement consistency through the correlation between scores on equivalent forms and consistency of scores across replications of a testing procedure (AERA, APA, NCME, 2014).

Reliability is necessary—though not sufficient—for the valid interpretation and use of scores. Test scores are considered reliable when an instrument produces similar scores across repeated administrations under certain conditions (e.g., short time intervals, no learning or intervention, varied items, no context effects, etc.). Because administering multiple forms at the same time or within a short period is often impractical, reliability is typically estimated using data from a single testing event.

Marginal reliability coefficients estimate the theoretical correlation between parallel forms of a test. The Classical Test Theory (CTT) definition of reliability is extended for Item Response Theory (IRT)-based computer adaptive tests (CAT), where students receive different item combinations and test only once per testing window.

The **conditional standard error of measurement (CSEM)** quantifies score imprecision at each scale score. Averaging CSEMs across a population yields the **standard error of measurement (SEM)**. Both metrics help interpret scores by indicating expected variability of scores across administrations. When reliability is high, standard error is low, meaning less variability of scores across administrations.

Marginal reliability and SEM analyses are conducted at the test level for the overall testing population and selected student groups. Group-level metrics provide evidence of fairness in an assessment.

Classification accuracy and consistency evaluate the reliability of placement decisions.

- **Classification accuracy** indicates the likelihood that students are placed in the correct placement level.
- **Classification consistency** indicates the likelihood that students would receive the *same* placement level across administrations, regardless of it being the correct placement.

Reliability analyses for *i-Ready Inform de lectura* use data from the fall 2025 administration (August to mid-October). Filtering rules retained as much valid data as possible, allowing one test per student for the testing window.

This chapter is organized as follows:

6. test-level reliability results for the overall population (marginal reliability and SEM)
7. results for selected student groups
8. classification accuracy and consistency estimates

6.3. Test-Level Reliability

Test-level measurement precision is evaluated using marginal reliability coefficients and the standard error of measurement (SEM), which is closely related to marginal reliability. The conditional standard error of measurement (CSEM) is the standard error of measurement at a specific scale score. This section concludes with plots of students' CSEMs against their scale scores, illustrating low measurement error across a wide range of proficiency levels. This precision reflects the effectiveness of our CAT algorithm and the quality and size of our item bank (see Chapter 4 for more information).

6.3.1. Marginal Reliability

Traditional reliability coefficients, such as coefficient alpha, reflect the precision of raw test scores (i.e., number correct). In contrast, marginal reliability reflects the precision of estimated latent ability, which is more appropriate for computer adaptive tests (CAT) because there are no fixed test forms, and reported scores are based on proficiency estimates.

The calculation of marginal reliability is rooted in the classical test theory (CTT) framework, where reliability is conceptualized as the squared correlation between two parallel test forms (Lord & Novick, 1968). Under CTT, an observed test score X is modeled as the sum of the true score T and an error term E . Using this definition, the reliability formula can be rearranged and expressed using quantities automatically produced for IRT-based tests (Sireci, Thissen, & Wainer, 1991), such as *i-Ready Inform de lectura*.

The reliability coefficient ($\hat{\rho}$) can therefore be written as:

$$\hat{\rho} = 1 - \frac{\sigma_E^2}{\sigma_\theta^2}, \text{ where} \quad (6.1)$$

$$\sigma_E^2 = \widehat{SEM}_\theta^2 = \frac{\sum_{i=1}^N CSEM_{\hat{\theta}_i}^2}{N}, \text{ and} \quad (6.2)$$

$$CSEM_{\hat{\theta}_i} = \frac{1}{\sqrt{I(\hat{\theta}_i)}}, \text{ and} \quad (6.3)$$

$$I(\hat{\theta}_i) = \sum_{j=1}^J P_{ij} (1 - P_{ij}), \quad (6.4)$$

and where i indexes students from 1 to N , j indexes items from 1 to J , σ_E^2 is variance of the error terms, σ_θ^2 is the variance of the observed scores on the logit scale, $\hat{\theta}_i$ is student i 's estimated proficiency on the logit scale, $CSEM_{\hat{\theta}_i}$ is student i 's CSEM for $\hat{\theta}_i$, $SEM_{(\theta)}$ is the SEM in θ or logit units (not for a particular value of θ), $I(\hat{\theta}_i)$ is the test information function for student i conditional on their proficiency estimate when using a Rasch IRT model and maximum likelihood scoring, and P_{ij} is the probability of a correct response of student i to item j , which is defined in Chapter 5.

The summand in Equation 6.4, $P_{ij}(1 - P_{ij})$, represents the information for a single item for a given student.

A reliability coefficient ranges from 0 to 1, with higher values indicating more reliable scores and lower SEMs. A common guideline is that reliability coefficients of .80 or higher are acceptable for tests of moderate lengths.

Marginal reliability coefficients are shown in **Error! Reference source not found.**, ranging from .90 to .95 indicating strong measurement precision across grades.

6.3.2. Standard Error of Measurement

An observed test score is not a student's *true* score. The true score is a theoretical concept and cannot be measured directly. However, through the careful construction of tests and rigorous psychometric practices, observed scores can estimate true scores with high precision. The uncertainty of these estimates is quantified by the standard error of the proficiency estimate, or the CSEM. The SEM is the average of all CSEMs across examinees (see Equation 6.2). This section focuses on CSEM.

Measurement precision can be improved by increasing test length, using a CAT algorithm, implementing strong item writing practices, and a robust item bank—all of which contribute to the strong measurement precision of *i-Ready Inform* de lectura.

Like reliability, the CSEM can be conceptualized in terms of repeated testing. If a student took the same test multiple times under identical conditions, their scores would vary, forming a distribution. The CSEM is the standard deviation of that distribution and is computed using Equation 6.3. Reliability and CSEM are inversely related: higher reliability corresponds to lower CSEM. Because this distribution is approximately normal, we can interpret CSEM units accordingly:

- about 68 percent confidence that the true score lies within ± 1 CSEM of the proficiency estimate
- about 95 percent confidence that the true score lies within ± 2 CSEMs

CSEMs are calculated for each student based on the items they received and vary in magnitude across the range of proficiency estimates—typically smaller near the center and larger at the tails—a common and expected pattern.

The computation of the CSEM (Equation 6.3) occurs on the underlying proficiency scale (the θ scale or the logit scale). Reported scores use a linear transformation with rounding and truncation rules to transform proficiency estimates and CSEMs from the underlying scale to the reporting scale of 200 to 500. The transformation formula is:

$$CSEM_{SS} = B \times CSEM_{\theta}, \quad (6.5)$$

where B is the scale transformation constant (34.4475 for Spanish Reading; see Chapter 5). Transformed CSEMs depend on the characteristics of the reporting scale, particularly its range: wider scales produce larger transformed CSEMs and SEMs, though this difference is purely an artifact of the transformation.

Transformed CSEMs can be averaged using Equation 6.2, with the θ subscripts replaced with SS to denote scale score to produce the SEM on the reporting score scale.

Alternatively, the SEM can be transformed directly using the equation:

$$SEM_{SS} = B \times SEM_{\theta}. \quad (6.6)$$

Because SEMs depend on scale characteristics, they cannot be directly compared across assessment programs. Scale-free metrics, such as reliability coefficients, are better suited for comparison.

Error! Reference source not found. presents the SEMs for this assessment range from 9.6 to 10.6. These small SEMs on a broad score scale indicate strong measurement precision across grades.

Table 6.1. Marginal Reliability and Standard Error of Measurement

Grade	Fall			Winter			Spring		
	N	Rel.	SEM	N	Rel.	SEM	N	Rel.	SEM
K	36,480	0.90	9.7	40,475	0.94	9.7			
1	46,125	0.95	9.6	44,538	0.95	9.7			
2	42,704	0.93	10.6	41,106	0.94	10.8			
3	37,389	0.94	10.1	36,760	0.94	10.2			
4	35,221	0.94	10.3	34,011	0.94	10.4			
5	30,552	0.94	10.4	30,357	0.94	10.5			
6	10,887	0.94	10.5	9,807	0.95	10.5			

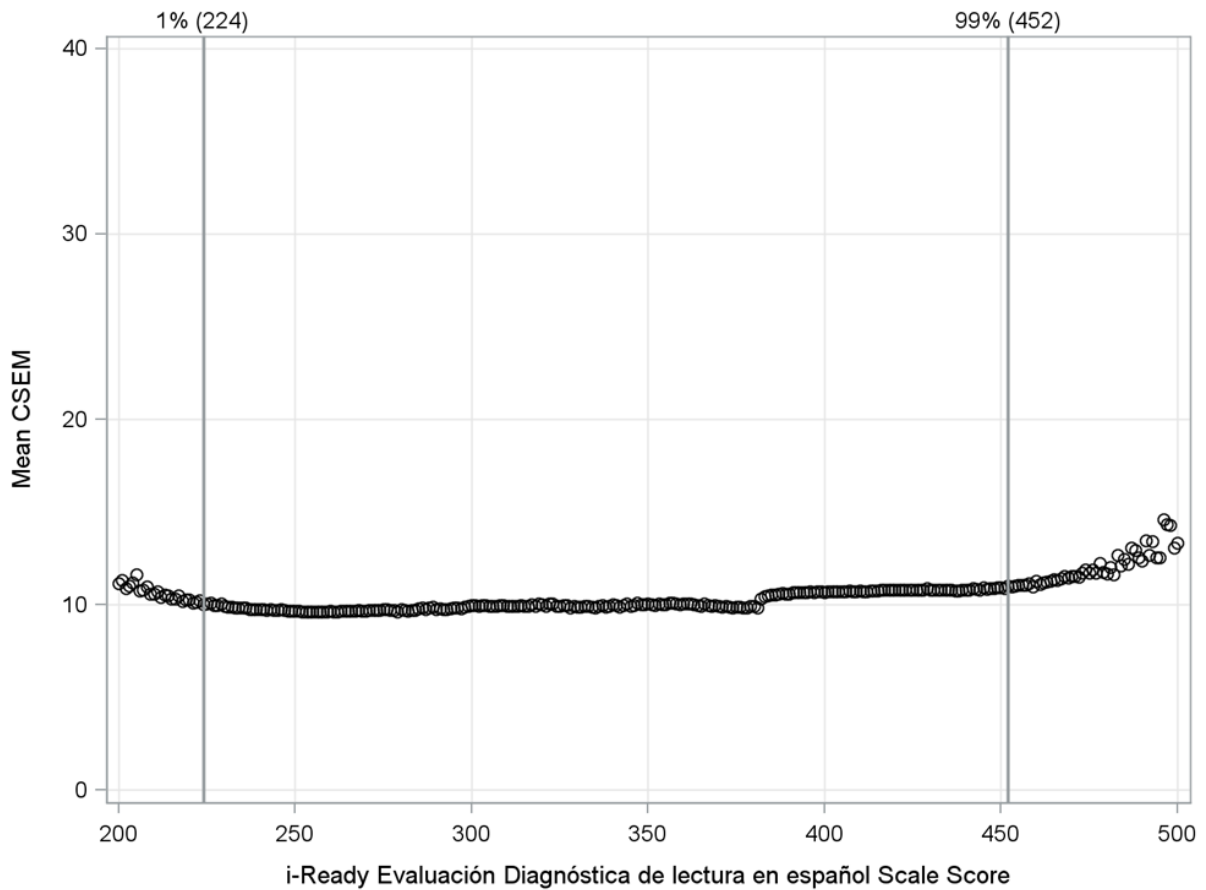


Figure 6.1 displays the mean CSEM at each scale score for students tested during the fall testing window (August to mid-October) across all grades. Vertical bars mark the 1st and 99th percentiles of the data used to generate the plot, illustrating that CSEM values are both low and relatively consistent across the vast majority of students.

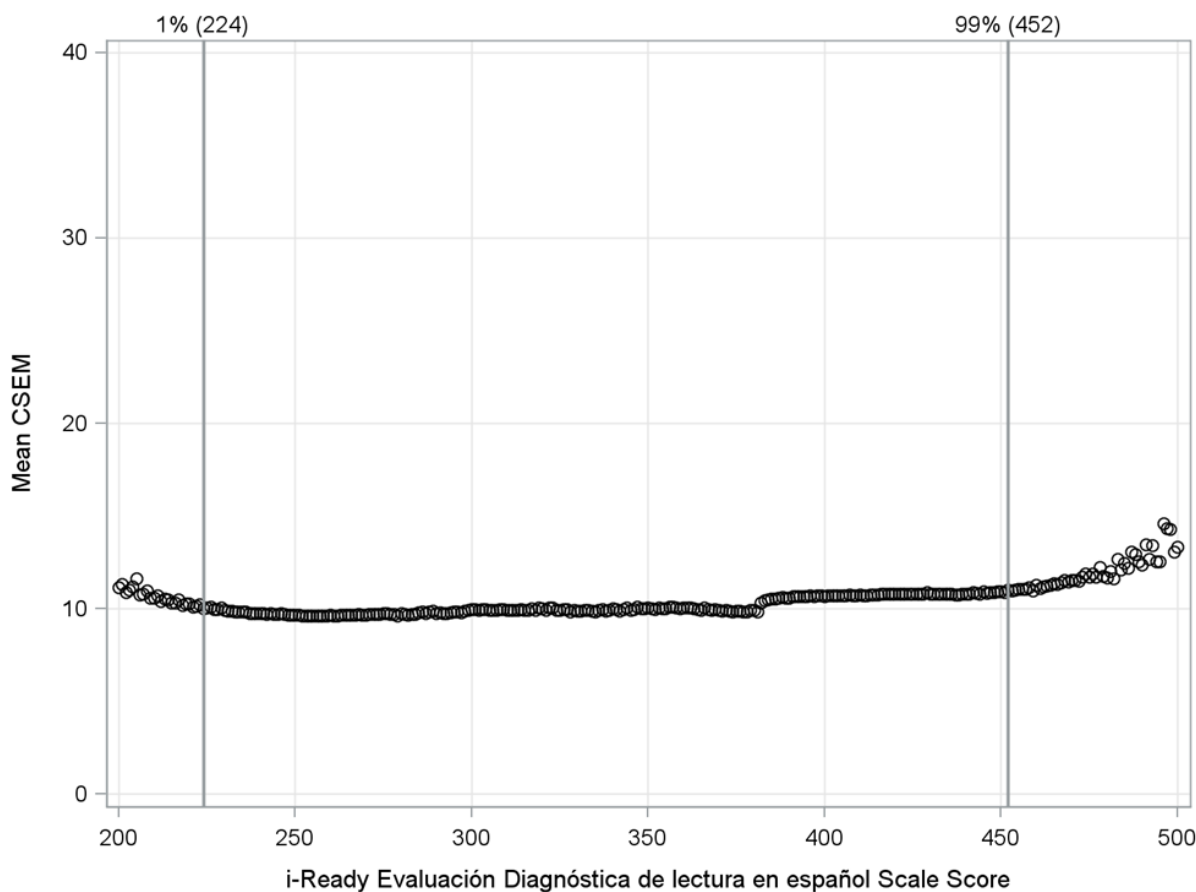


Figure 6.1. Conditional SEM of *i-Ready Inform de lectura*

6.4. Reliability for Select Groups

Although demographic data are limited and not provided for every student, it is important to evaluate measurement precision for select student groups where possible. Ideally, marginal reliability and SEM for these groups should be similar to the indices computed for all students. However, differences in proficiency distributions can affect these statistics—particularly marginal reliability, which depends on overall score variance.

Marginal reliability coefficients and SEMs were computed at the test level for select student groups and are presented in Table 6.2. Overall, reliability coefficients and SEMs for economically disadvantaged students, English Learners, students in special education, different racial/ethnic groups, and genders are similar to those for all test-takers, indicating strong and consistent measurement precision across groups.

Table 6.2. Marginal Reliability and Standard Error of Measurement for Select Groups

Grade	Fall			Winter			Spring		
	N	Rel.	SEM	N	Rel.	SEM	N	Rel.	SEM
Economically Disadvantaged									
K	12,393	0.90	9.7	15,566	0.94	9.7			
1	15,370	0.95	9.6	16,625	0.95	9.7			
2	14,247	0.93	10.6	15,200	0.93	10.8			
3	11,014	0.93	10.2	12,095	0.93	10.3			
4	9,893	0.93	10.4	10,897	0.93	10.5			
5	8,772	0.92	10.5	9,636	0.93	10.6			
6	2,240	0.94	10.5	1,779	0.94	10.6			
English Learners									
K	17,064	0.89	9.7	20,551	0.94	9.7			
1	22,361	0.95	9.6	22,769	0.95	9.7			
2	20,821	0.93	10.6	21,046	0.93	10.8			
3	17,373	0.94	10.1	17,508	0.93	10.3			
4	16,041	0.93	10.3	16,134	0.92	10.4			
5	13,312	0.93	10.4	13,362	0.93	10.5			
6	3,458	0.93	10.4	3,097	0.93	10.5			
Special Education									
K	1,885	0.88	9.7	2,813	0.93	9.8			
1	3,007	0.94	9.6	3,315	0.95	9.7			
2	3,070	0.93	10.4	3,263	0.94	10.6			
3	2,849	0.94	9.9	3,111	0.94	10			
4	2,705	0.95	10.0	3,016	0.94	10.1			
5	2,471	0.94	10.2	2,678	0.95	10.2			
6	753	0.95	10.3	676	0.96	10.3			
American Indian or Alaskan									
K	2,530	0.90	9.7	2,867	0.94	9.7			
1	2,987	0.95	9.6	3,208	0.95	9.7			
2	2,783	0.93	10.6	3,025	0.93	10.8			
3	1,941	0.93	10.1	2,060	0.93	10.3			
4	1,931	0.93	10.3	2,075	0.93	10.4			
5	2,033	0.92	10.4	2,271	0.92	10.5			
6	600	0.92	10.5	620	0.92	10.6			

Table 6.2. Marginal Reliability and Standard Error of Measurement for Select Groups

Grade	Fall			Winter			Spring		
	N	Rel.	SEM	N	Rel.	SEM	N	Rel.	SEM
Asian									
K	285	0.91	9.7	496	0.94	9.7			
1	513	0.95	9.6	541	0.95	9.7			
2	504	0.94	10.7	522	0.94	10.8			
3	500	0.95	10	507	0.95	10.2			
4	419	0.96	10.3	470	0.95	10.3			
5	349	0.95	10.3	349	0.96	10.4			
6	124	0.93	10.5	108	0.95	10.7			
Black or African American									
K	803	0.89	9.7	1,316	0.93	9.7			
1	1,376	0.95	9.6	1,540	0.95	9.7			
2	1,228	0.94	10.6	1,273	0.95	10.7			
3	1,176	0.95	10.0	1,279	0.95	10			
4	998	0.95	10.1	1,053	0.96	10.2			
5	858	0.95	10.2	933	0.96	10.2			
6	292	0.95	10.3	304	0.96	10.4			
Hispanic									
K	29,511	0.90	9.7	33,738	0.94	9.7			
1	37,580	0.95	9.6	37,662	0.95	9.7			
2	35,694	0.93	10.6	34,758	0.93	10.8			
3	31,161	0.94	10.1	31,040	0.93	10.2			
4	29,266	0.94	10.3	28,470	0.93	10.4			
5	25,748	0.93	10.4	25,662	0.93	10.5			
6	8,524	0.93	10.5	7,638	0.94	10.6			
Native Hawaiian or Pacific Islander									
K	158	0.90	9.7	197	0.93	9.7			
1	222	0.94	9.6	231	0.95	9.6			
2	222	0.93	10.5	236	0.94	10.7			
3	188	0.94	10	186	0.94	10.1			
4	168	0.95	10.2	191	0.95	10.2			
5	157	0.95	10.2	152	0.94	10.4			
6	**	**	**	**	**	**			

Table 6.2. Marginal Reliability and Standard Error of Measurement for Select Groups

Grade	Fall			Winter			Spring		
	N	Rel.	SEM	N	Rel.	SEM	N	Rel.	SEM
Other Race									
K	**	**	**	**	**	**			
1	**	**	**	**	**	**			
2	**	**	**	**	**	**			
3	**	**	**	**	**	**			
4	**	**	**	**	**	**			
5	**	**	**	**	**	**			
6	**	**	**	**	**	**			
Two or More Races									
K	3,589	0.90	9.7	3,064	0.94	9.7			
1	4,015	0.95	9.6	3,217	0.95	9.7			
2	3,782	0.93	10.6	2,656	0.94	10.8			
3	3,032	0.94	10.2	2,227	0.94	10.3			
4	2,866	0.93	10.3	2,037	0.93	10.4			
5	2,402	0.92	10.4	1,798	0.93	10.5			
6	468	0.95	10.5	388	0.95	10.5			
White or Caucasian									
K	21,568	0.90	9.7	24,159	0.94	9.7			
1	24,876	0.95	9.6	26,026	0.95	9.7			
2	23,195	0.93	10.6	23,743	0.93	10.8			
3	20,588	0.94	10.1	21,322	0.93	10.3			
4	19,716	0.94	10.3	19,680	0.93	10.4			
5	16,832	0.93	10.4	17,561	0.94	10.5			
6	3,792	0.93	10.5	3,461	0.94	10.6			
Male									
K	16,238	0.90	9.7	18,806	0.94	9.7			
1	20,872	0.95	9.6	21,021	0.95	9.7			
2	19,881	0.94	10.6	19,318	0.94	10.8			
3	17,640	0.94	10.1	17,500	0.94	10.2			
4	16,446	0.94	10.2	16,009	0.94	10.4			
5	14,285	0.94	10.3	14,234	0.94	10.5			
6	4,986	0.94	10.4	4,436	0.95	10.5			

Table 6.2. Marginal Reliability and Standard Error of Measurement for Select Groups									
Grade	Fall			Winter			Spring		
	N	Rel.	SEM	N	Rel.	SEM	N	Rel.	SEM
Female									
K	17,198	0.90	9.7	19,994	0.93	9.7			
1	21,563	0.94	9.6	21,653	0.95	9.7			
2	20,712	0.93	10.6	20,298	0.93	10.8			
3	18,029	0.94	10.1	17,907	0.93	10.3			
4	17,158	0.93	10.3	16,714	0.93	10.4			
5	14,892	0.93	10.4	14,898	0.93	10.5			
6	5,170	0.93	10.5	4,640	0.94	10.6			

Note: Rel. = Reliability

6.5. Reliability of Classification

Classification accuracy and consistency metrics describe the reliability of assigning students to placement levels using one or more cut scores. These metrics are influenced by factors such as the number and location of the cut scores, the distribution of student proficiency, and the CSEM of proficiency estimates (which depends on test length).

- **Classification accuracy** is the degree to which students' assigned placement levels match their *true* placement levels.
- **Classification consistency** is the degree to which students would be placed in the same placement level across administrations, regardless of it being the correct placement.

Both metrics are expressed as proportions. Because classification consistency reflects two sources of error (two administrations), it is always lower than classification accuracy, which reflects one source of error.

We estimated classification accuracy using Rudner's (2000, 2005) method, which applies key IRT results: $\hat{\theta}_i$ follows a normal distribution centered at θ_i (i.e., $\hat{\theta}_i$ is an unbiased estimator of θ_i) with a standard deviation equal to $CSEM_{\hat{\theta}_i}$.

Using these properties, we constructed a normal distribution for each student based on their $\hat{\theta}_i$ and $CSEM_{\hat{\theta}_i}$, overlaid placement thresholds onto that normal distribution, and computed the area under the curve for each placement level.

The student's *true* placement level is assumed to be the one in which the student's actual score is located. The proportion under the normal curve associated with this placement level is the

classification accuracy for student i . These student-level values are averaged to produce the test-level classification accuracy statistic.

Wyse and Hao (2012) extended Rudner's (2000, 2005) method to compute classification consistency from the probability by placement-level vector described above. Each element of the probability-by-level vector is squared to represent the likelihood of being classified in the same level twice. The sum of these squared probabilities for each student yields a consistency metric, which is then averaged across students.

Classification accuracy and consistency results for August to mid-October of the fall 2025 administration appear in Table 6.3. Columns labeled "Two Levels Below," "One Level Below," "Early On Grade," and "Mid On Grade" correspond to the cut scores for those placement levels. Each column shows classification accuracy or consistency relative to that cut score. For example, the "One Level Below" column partitions students into two categories: two or more levels below and one level below and above.

The "Three or More Levels Below" level does not exist for Grades K–2, and the "Two Levels Below" level does not exist for Grade K. Therefore, these grade and placement level combinations are not reported. The "All" columns use all four cut scores, dividing students into up to five groups.

Results of these analyses are:

- Classification accuracy for single cut scores is $\geq .90$.
- Classification consistency is slightly lower than classification accuracy, as expected.
- Classification accuracy for all cuts combined is $\geq .75$.
- Classification consistency for all cuts is $\geq .67$.

Table 6.3. Classification Accuracy and Consistency											
Grade	N	Classification Accuracy					Classification Consistency				
		Two Levels Below	One Level Below	Early On Grade	Mid On Grade	All	Two Levels Below	One Level Below	Early On Grade	Mid On Grade	All
K	36,480	—	—	0.92	0.98	0.89	—	—	0.88	0.97	0.85
1	46,125	—	0.95	0.92	0.97	0.83	—	0.93	0.89	0.95	0.77
2	42,704	—	0.90	0.93	0.95	0.78	—	0.87	0.91	0.94	0.71
3	37,389	0.95	0.93	0.93	0.96	0.76	0.93	0.90	0.90	0.94	0.67
4	35,221	0.95	0.92	0.91	0.98	0.76	0.92	0.89	0.88	0.97	0.67
5	30,552	0.93	0.90	0.94	0.97	0.75	0.89	0.87	0.92	0.96	0.67
6	10,887	0.91	0.94	0.95	0.98	0.79	0.87	0.91	0.93	0.97	0.72

6.6. Discussion

The analyses and results presented in this chapter demonstrate that the depth of the item bank, the effectiveness of the item selection algorithm, and the design of the test flows collectively produce reliable overall score estimates across the proficiency range.

Although demographic data are limited, group-level reliability results indicate that scores for each group are estimated with a degree of precision comparable to that of the overall population.

Classification accuracy and consistency coefficients reinforce these findings of the reliability of *i-Ready Inform de lectura*.

Chapter 7. Setting Standards

7.1. Chapter Summary

This chapter explains how score categories, called placement levels, were established for *i-Ready Inform™ de lectura* so educators can more meaningfully interpret student performance in Spanish reading. The chapter walks through the standard-setting process, explaining how experts and stakeholders worked together to define what it means for students to be at different placement levels. It describes the use of Embedded Standard Setting (ESS) methodology (Lewis & Cook, 2020), where subject matter experts align test items to detailed placement-level descriptions and then use statistical analyses to set cut scores that separate the placement levels. The chapter also covers how policy decisions and empirical data were combined to finalize the cut scores for each placement level, ensuring that the categories reflect student achievement and support instructional decisions.

7.2. Introduction

7.2.1. The Basis for Standard Setting

i-Ready Inform de lectura is a valuable tool for educators because it provides criterion-referenced placement levels that inform many components of the score reports. These placement levels are aligned with grade-level expectations for student performance in Spanish reading. Through the standard setting process, we link these expectations with student achievement on the assessment.

The Standards for Educational and Psychological Testing (AERA et al., 2014) provide guidelines for setting cut scores and documenting procedural evidence to support the validity argument for score interpretations and uses. Specifically, Standard 5.21 states: “When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly” (p. 107).

For a criterion-referenced assessment like *i-Ready Inform de lectura*, the overall validity argument relies on the use of a procedure that links descriptions of student performance to the test’s score scale. This ensures that placement levels are meaningful and not arbitrary (Kane, 2013). Therefore, appropriate methodologies must be used to determine cut scores. This chapter summarizes the standard-setting procedures and statistical methods used to determine cut scores for *i-Ready Inform de lectura*.

7.2.2. Overview of Method

There are several standard setting methodologies that align with the guidelines established by the Standards for Educational and Psychological Testing (AERA et al., 2014). The selection of a methodology generally depends on the test item types, the characteristics of the assessment, and how well the resulting evidence supports the overall validity argument (Hambleton & Pitoniak, 2006).

This chapter provides procedural evidence from the standard setting of *i-Ready Inform de lectura* to support the overall validity argument.

There are five key placement levels, both overall and for the domains: Below Grade Level, Early Grade Level, Mid Grade Level, Late Grade Level, and Above Grade Level. The labels indicate that Below and Above Grade Level placements are considered *off-grade* level placements. In contrast, Early, Mid, and Late Grade Level placements are considered *on-grade* level placements. Below and Above Grade Level placements can be further delineated by the number of grade levels below or above.

Mid and Late cuts were established using industry-standard Embedded Standard Setting (ESS) methodology (Lewis & Cook, 2020). This methodology involves aligning items to Placement Level Descriptors (PLDs) by Subject Matter Experts (SME) and conducting ESS analyses to compare empirical data with those professional judgment alignments. Curriculum Associates contracted with Creative Measurement Solutions to apply the ESS methodology, which estimates cut scores by optimizing the alignment between the SME item-PLD alignments and empirical data. The Early cuts were determined by Curriculum Associates staff and external stakeholders through policy-based decisions that differentiates below grade level from on grade level. Cut scores for the Below Grade Level and Above Grade Level categories were determined based on heuristics that align to adjacent grade levels and consider the measurement error of the test. Cut scores for all placement levels, both overall and by domain, can be found in the *i-Ready Inform de lectura* document.

7.3. Placement Levels

7.3.1. Goals of Setting Placement Levels

An essential step in developing an assessment program is determining the minimum level of student proficiency required for classification into each placement level. These classifications simplify the interpretation of the score scale by grouping scores into distinct categories that provide meaningful descriptions of student performance.

For *i-Ready Inform de lectura*, placement levels are defined by detailed descriptions of the performance required within each score range. These descriptions, referred to as PLDs, were developed by subject-matter experts, making this assessment criterion referenced.

7.3.2. Placement-Level Descriptors

7.3.2.1 Policy-Level PLDs

Policy-level PLDs are broad, descriptive statements that generally outline the progression of skills and knowledge students acquire through learning (Egan et al., 2012). These statements concisely describe student performance in relation to the expectations specified in the College and Career Readiness Standards for each grade level and subject.

For *i-Ready Inform de lectura*, policy-level PLDs also provide recommendations for the most appropriate level of instruction for students assigned to each placement level. The PLDs are grounded in sets of state Spanish Reading standards, such as those used in Texas and California. They reflect the content that is taught in classes to ensure students attain proficiency in the standards, and they articulate the skills that guide item development.

The five policy-level PLDs are shown in **Error! Reference source not found.**, which is the same as **Error! Reference source not found.** and is repeated here for ease of reference.

Table 7.1. Policy-Level PLDs	
Placement Level	Description
Below Grade Level	Remediation focused on below-grade level material is recommended to help fill in gaps in students' foundational knowledge. Students in this level are not close to meeting the expectations of College and Career Readiness Standards for their grade level.
Early Grade Level	Students in this level will benefit from on-grade level instruction to help them meet the expectations of College and Career Readiness Standards for their grade level. Students in the early on grade level have only partially met these grade-level expectations.
Mid Grade Level	Students in this level will benefit from instruction in late on-grade level topics. These students have met the minimum requirements for the expectations of College and Career Readiness Standards in their grade level.
Late Grade Level	Students in this level will benefit from late on-grade level enrichment and will be ready for instruction focused on topics typically covered in the beginning of the subsequent grade level. Students in the late grade level have successfully met or surpassed the grade-level expectations of College and Career Readiness Standards.
Above Grade Level	Students in this category will benefit from above-grade level instruction. Students in the above grade level have successfully met or surpassed all the expectations of College and Career Readiness Standards for their grade level as well as some expectations from subsequent grade levels.

7.3.2.2 Range PLDs

To help inform instructional recommendations and placement decisions, SMEs and assessment experts created Range PLDs for standard setting. These PLDs provide specific descriptions of the knowledge and skills expected of students at each placement level, tailored to each grade and domain. Written in Spanish, these PLDs cater specifically to Spanish reading content. The development of these Range PLDs considered several factors, including various sets of state Spanish Reading Standards.

As detailed in Chapter 2, Range PLDs include anchor claims, which provide detailed statements explaining expectations for students in relation to the summary claim at each placement level. Each test item is mapped to a single anchor claim, and each anchor claim corresponds to one summary claim and placement level combination. Thus, anchor claims specify the evidence required for

students to demonstrate their knowledge and skills as outlined in the summary claim at the three on-grade placement levels: Early, Mid, and Late, to be considered proficient in Spanish reading overall or in the domain at that grade level.

7.3.2.3 Item-PLD Alignment

SMEs aligned each item to a placement level using the Range PLDs. SMEs identified the most relevant summary claim for each item, ensuring the item showed evidence that the student had acquired the necessary skills and knowledge for proficiency related to that summary claim. Within the summary claim, SMEs pinpointed the more specific skill or placement level (anchor claim) that the item targeted.

After this initial alignment, SMEs calibrated their alignment judgment on a sample set of alignments for each grade and domain, cross-checking each other's work and consulting with an assessment development expert. When sample sets from each domain were completed, SMEs discussed their alignments and incorporated feedback to improve future alignments. SMEs met periodically to review alignments and address any issues. Occasionally, with the assessment development expert's approval, one or more claims in the PLDs were edited to better reflect the content of the test items.

7.4. Methods and Procedures

7.4.1. Embedded Standard Setting Method

The Embedded Standard Setting (ESS) approach establishes cut scores by optimizing the consistency between professional judgments of item–PLD alignment and empirical data (Lewis & Cook, 2020). ESS requires several key components (see Figure 7.1): well-articulated PLDs, assessment items that align to the PLDs, and empirical data that corroborate the item–PLD alignments. ESS was utilized to determine the Mid and Late cut scores, which differentiate the three on-grade placement levels: Early, Mid, and Late.

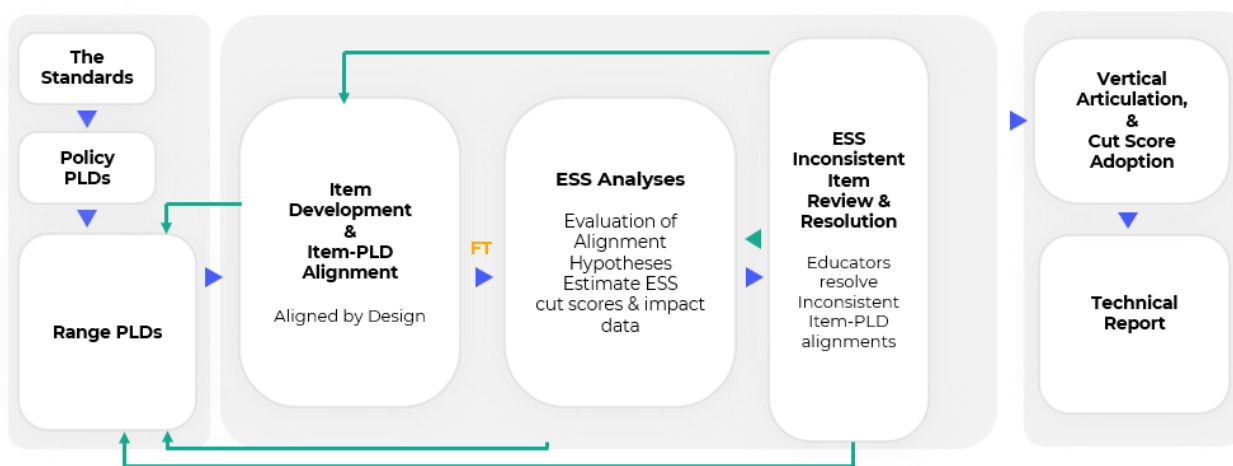


Figure 7.1. Embedded Standard Setting Process (Source: Dan Lewis, Creative Measurement Solutions)

Two rounds of ESS Analyses were conducted with a review of ESS inconsistent items occurring in between and an additional review of ESS inconsistent items following the second round.

7.4.1.1 Empirical Data

Existing items from the previous Assessment of Spanish Reading (ASR) were evaluated, and if appropriate, used for *i-Ready Inform de lectura*. Operational data for these items were considered for the standard setting process. Additionally, field test data collected from spring 2023 to spring 2024 were used to estimate Item Response Theory (IRT) item parameters for use in the ESS analyses. An additional pilot study was conducted in fall 2024⁵ using an adaptive testing platform to obtain student responses (and test scores). This study provided impact data to evaluate the percentage of students classified into each placement level based on the determined cut scores.

The selection of an IRT response probability is a policy decision guided by the language used in the PLDs. The two most frequently used response probability values are RP50 and RP67. RP50 indicates a lower performance expectation than RP67. Specifically, at RP50, a student with this proficiency is expected to have a 50% chance of correctly answering an item at the cut score for that level. In contrast, with RP67, a student with this proficiency is expected to have a 67% chance of correctly answering an item at the cut score of that level. RP67 is typically chosen when PLDs describe knowledge, skills, and other attributes that students in each level are expected to demonstrate. Therefore, RP67 was adopted as the response probability value for determining cut scores for *i-Ready Inform de lectura*. This approach is consistent with other *i-Ready Assessments*.

7.4.1.2 ESS Analyses

Given the item-PLD alignments and empirical item difficulty estimates, the ESS software (EmStanS) calculates the cut score between adjacent placement levels (early/mid; mid/late) using one of several different optimization functions. For *i-Ready Inform de lectura*, we used the ESS-Weight (Lewis, Lee, & Choi, 2021) optimization function.

ESS-Weight can be expressed mathematically as:

$$\text{ESS-Weight} \equiv \arg \min_c \sum_{i=1}^n I(\text{d-inconsistent}) \cdot |D_i| , \quad (1)$$

$$D_i \equiv RP_i - c , \quad (2)$$

⁵ The pilot included items from grades K-5 taken by students in grades 1-6. Grade 6 items were not included in the pilot. Grade 6 impact data will be evaluated after the first operational year of data collection in 2025-2026.

where RP_i is the IRT RP location for item i and c is a cut score candidate. An item is considered ESS-Inconsistent relative to a cut score (or a cut score candidate) when empirical data are not consistent with the SME Item-PLD Alignment. Given a candidate cut score, the Distance, D_i , of an ESS-Inconsistent item, i , is defined as the minimum number of scale score points that the item's IRT location must shift to place the item at a border of the Item-PLD Aligned level. In this notation the second term in the summand, $|D_i|$, the expression for ESS-Distance, is clearly seen as the weight for item i , $W_i = |D_i|$.

An item is ESS-Inconsistent relative to the Mid cut score if either of the following conditions is met:

- The item is aligned to the Early level, but the item's IRT RP location is at or above the Mid cut score.
- The item is aligned to the Mid level, but the item's IRT RP location is below the Mid cut score.

Similarly, an item is ESS-Inconsistent relative to the Late cut score if either of the following conditions is met:

- The item is aligned to the Early or Mid level, but the item's IRT RP location is at or above the Late cut score.
- The item is aligned to the Late level, but the item's IRT RP location is below the Late cut score.

ESS cut scores are estimated by identifying the minimum value of ESS-Weight for all cut score candidates across the test scale.

7.4.1.3 ESS-Inconsistent Item Review and Resolution

The ESS-Inconsistent item review and resolution process, as described by Lewis & Cook (2020) and Brice & Mix (2021), involves SMEs examining items with hypothesized Item-PLD alignments that do not align with empirical data. The goal is to understand the cause of the Inconsistency and to address it. Creative Measurement Solutions provided Curriculum Associates with information on these ESS-inconsistent items to facilitate their review and resolution.

The ESS-Inconsistent item review team comprised SMEs who had previously reviewed the item-PLD alignments. Each SME involved in the review was a Spanish speaker and also an expert in ELA content and assessment. They received the necessary documentation and were trained in the review process and its purpose by a Curriculum Associates assessment design expert. This training focused on identifying the likely source of item inconsistencies and recommending possible resolution actions.

Possible actions to resolve inconsistent items included:

Outcome 1: Add one or more new anchor claims to better reflect the content of the items. This is a claim-level recommendation.

Outcome 2: Combine anchor claims or shift anchor claims to a different placement level. This is a claim-level recommendation.

Outcome 3: Realign one or more items to an updated or different placement level descriptor. This is an item-level recommendation.

Outcome 4: Suggest one or more item edits, which may include disabling the item and removing it from the operational bank. This is an item-level recommendation.

Outcome 5: Make no changes; accept the inconsistency. This would be indicated when SME judgment resulted in no content-based explanation for a discrepancy in SME-aligned and empirical placement levels.

During training, SMEs were guided through the process of viewing items and documenting their decision making, which was discussed and demonstrated thoroughly.

For the anchor claim review and resolution, 69 anchor claims, 52 testlets, and 452 items were reviewed. The SMEs recommended updates to some PLDs and edits to some anchor claims in grades K, 1, 2, 3, and 4. Before finalizing these recommended updates, Curriculum Associates performed a content-based vertical articulation. This process involved inspecting the recommended PLD changes across grades to ensure consistency and parallel structure of the updated claims. An assessment development expert reviewed and accepted all recommended updates to PLDs, resulting in revisions to 36 of the 69 reviewed anchor claims. Additionally, at the item level, the expert reviewed and accepted all proposed changes to SME-aligned levels.

This process led to updated placement levels for 220 of 263 items in round 1 and 173 of 189 items in round 2.

Empirical data with the new item-PLD alignments were submitted to Creative Measurement Solutions to rerun the ESS Analyses and update cut score recommendations for the Mid and Late cuts. Mid cuts were established for grades K-6 and Late cuts were established for grades K-5. Grade 6 impact data will be evaluated after the first operational year of data collection before establishing a Late cut for grade 6.

7.5. Extensions

7.5.1. Early Cut Scores

i-Ready Inform de lectura categorizes student placement into three on-grade placement levels per grade: Early, Mid, and Late, as differentiated by the ESS Mid and Late cut scores. The Early cut is the threshold determining whether students are classified as below grade level rather than on grade level. The Early cut serves as the starting point for all other below grade level cuts and is determined by a

policy-based Early cut score. Various policy-based approaches were considered to establish the Early cut score that distinguishes the Below level from the Early level, including:

1. Application of a heuristic to the ESS cuts. The cut is determined in reference to student performance at the next lower grade. This approach resulted in the highest cut scores and is, therefore, the most conservative approach.
2. Application of a heuristic to the ESS cuts minus 1 SEM. This approach is a less conservative and resulted in lower cut scores.
3. Use of the Assessment of Spanish Reading (ASR) Partially Met Cuts. This approach resulted in the lowest cut scores and is the most lenient approach.

Impact data from these three approaches were presented to seven stakeholder groups, which included educators, school administrators, and Curriculum Associates staff who work directly with schools. Based on stakeholder feedback, the Early cut was established by using a combination of approaches 1 and 2. The heuristic was applied to obtain Early cuts for Grades K–1, and the heuristic minus 1 SEM was applied to obtain the Early cuts for Grades 2–6.

Stakeholders recommended against using the previous Assessment of Spanish Reading (ASR) as a basis for the Early cut (Option 3) because *i-Ready Inform de lectura* differs significantly from ASR. For Grades K–1, a direct application of the heuristic (Option 1) was used to set the Early cut. This decision was influenced by stakeholders' recommendation for a more conservative cut score at these lower grade levels to ensure early intervention for students who need instructional support. For the Early cut at Grades 2–6, the application of the heuristic was adjusted by subtracting 1 SEM (Option 2), as stakeholders believed that a slightly less conservative cut was suitable for these higher grade levels. This adjustment produced impact data that aligned with educators' expectations.

7.5.2. Above and Below Grade Cut Scores

Students may receive grade level placements above or below their chronological grade. After establishing the Early, Mid, and Late cut scores, the Above and Below Grade cuts were systematically set using a heuristic approach. For each grade, the Above Grade cut was determined based on the Late Grade cut for the next higher grade level. Conversely, the Below Grade cut was set in reference to the Early Grade cut for the grade level immediately lower. There are no restrictions on Below Grade placements, allowing students to receive a placement level at any grade level below their chronological grade. However, Above Grade placements are limited to two grade levels higher. The stakeholder groups involved in setting the Early cuts also reviewed the impact data related to the above and below grade cut scores.

7.5.3. Domain Cut Scores

The process for establishing the overall cut scores for *i-Ready Inform de lectura* involved utilizing item information for the overall test to maximize the data evaluated during the standard setting process. To

further maximize the data informing the cut scores, a top-down approach was taken in which the overall cuts were applied down to each domain. Curriculum Associates SMEs reviewed the percentage of students within each placement level across domains to ensure that using the overall cuts as domain cuts was pedagogically sound. This consistent application of cut scores across domains enables meaningful interpretation and comparisons of scores both within and across domains and between domain scores and the overall score.

7.5.4. Finalizing the Cut Scores

An internal Curriculum Associates team considered several factors when determining the final cut scores:

- the percentages of students from the pilot study who placed above or below the established cut scores
- the statistical relationship with results from the Texas and California state assessments
- the progression of the cut scores across grade levels

Vertical articulation was conducted to ensure the reasonableness of cut scores and the corresponding impact data across grade levels. Table 7.2 lists the final 2025-2026 on-grade-level scale score cuts for the overall test and each domain. Table 7.3 through Table 7.7 display the pilot study impact data corresponding to these adopted, vertically articulated cut scores.

Table 7.2. 2025-2026 Cut Scores (Common for Overall and each Domain)			
	Cut Scores		
Grade	Early*	Mid	Late
K	278	322	370
1	322	366	412
2	356	392	426
3	382	414	448
4	404	437	464
5	427	449	480
6	438	461	n/a

* Note: The Early cut scores were established by policy; Mid and Late cuts were set using ESS.

Table 7.3. Overall Impact Data				
	Percent in Level			
Grade	Below	Early	Mid	Late
K	11.0%	37.4%	42.2%	9.4%

Table 7.3. Overall Impact Data				
	Percent in Level			
Grade	Below	Early	Mid	Late
1	24.7%	51.1%	23.4%	0.8%
2	42.4%	38.0%	17.9%	1.7%
3	54.3%	33.7%	10.5%	1.5%
4	61.6%	25.3%	10.7%	2.5%
5	67.3%	16.4%	11.7%	4.7%
6	*	*	*	*

*No impact data were available from the fall pilot study for grade 6.

Table 7.4. Vocabulary Impact Data				
	Percent in Level			
Grade	Below	Early	Mid	Late
K	11.4%	36.3%	38.6%	13.7%
1	31.4%	42.9%	22.1%	3.7%
2	43.0%	35.3%	16.9%	4.8%
3	48.7%	31.2%	17.9%	2.2%
4	51.0%	30.6%	15.5%	2.9%
5	60.4%	21.9%	13.3%	4.3%
6	*	*	*	*

*No impact data were available from the fall pilot study for grade 6.

Table 7.5. Comprehension Impact Data				
	Percent in Level			
Grade	Below	Early	Mid	Late
K	18.1%	40.1%	34.6%	7.2%
1	41.3%	37.8%	19.9%	1.0%
2	47.4%	32.8%	17.2%	2.5%
3	61.4%	26.9%	10.1%	1.7%
4	65.5%	22.3%	9.6%	2.6%
5	69.1%	13.7%	12.4%	4.9%
6	*	*	*	*

*No impact data were available from the fall pilot study for grade 6.

Table 7.6. Phonological Awareness Impact Data				
	Percent in Level			
Grade	Below	Early	Mid	Late
K	13.1%	25.8%	34.8%	26.3%
1	20.5%	31.5%	34.6%	13.4%

Table 7.7. Phonics Impact Data				
	Percent in Level			
Grade	Below	Early	Mid	Late
K	10.8%	31.0%	47.0%	11.3%
1	25.1%	45.4%	26.4%	3.1%
2	32.4%	41.2%	20.3%	6.1%
3	*	*	*	*

*No impact data were available from the fall pilot study for grade 3 Phonics.

Chapter 8. Score Reports

8.1. Chapter Summary

Score reports are essential components of the assessment system because they communicate the meaning and intended use of scores to stakeholders who support learning. This chapter introduces the reports for *i-Ready Inform de lectura* available to teachers and administrators in the online reporting system. These reports provide descriptive information about students' current proficiency levels and prescriptive guidance to enhance learning. This chapter also identifies the interpretive guidance included within the reports and external resources that help users draw valid inferences from assessment results.

8.2. Introduction

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) require test developers to provide clear explanations of intended score uses and interpretations in all reporting practices. Standard 5.1 emphasizes the importance of explaining test scores to users, including the “characteristics, meaning, and intended interpretation of scale scores, as well as their limitations.”

8.2.1. Levels of Reporting

The *i-Ready* reporting platform provides practical, intuitive reports that educators can access in real time through their *i-Ready* dashboard. Reports can be printed, though all report images in this report reflect their appearance online. Data can also be exported as .csv files.

Reports are available—or scheduled to become available—at the following levels:

- **Student-level reports** (available starting 2025-2026) primarily communicate individual student results to teachers, though administrators may also access them. These reports include detailed information and clickable features to enhance understanding. They provide clear descriptions of scores, placement levels, and other achievement indicators. The *For Families* report provides student results in lay terms to support teacher–family conversations and identify appropriate instructional activities to do at home.
- **Class-level reports** (available starting 2025-2026) help teachers and other administrators understand class-level results. These reports aggregate student-level data and identify instructional resources tailored to class results.
- **School- and district-level reports** (targeted for 2026-2027) enable administrators to review aggregated results for grades, schools, or districts. These reports support instructional and resource planning and link to instructional resources focused on selected skills.

8.3. *i-Ready Inform* Results Reports (Spanish Reading)

8.3.1. Student Level

The student-level report provides criterion-referenced information about a student's performance in Spanish reading overall and by domain. Delivered online, the report is dynamic; teachers can easily navigate across domains or administrations for a student or switch between students using a dropdown menu. The report includes skill-based information to inform instruction.

8.3.1.1 Purpose of the Report

After a student completes *i-Ready Inform de lectura*, the *i-Ready Inform* Results Report provides scores, placements, and detailed insights into performance. This report is available at the student, class, and school/district levels.

- At the **student level**, the report provides a comprehensive view of a student's performance for a given administration.
- At the **class level** (available starting 2025-2026) and **school/district level** (targeted 2026-2027), the report is intended for teachers and administrators and summarizes the distribution of overall and domain placements.

Students can view their scale scores for each completed assessment in their *i-Ready* dashboard, but they do not see their placement level. Curriculum Associates intends for teachers to share assessment results with students (see the Theory of Action in Chapter 1).

8.3.1.2 Overall Scale Score

The overall scale score quantifies student proficiency in Spanish reading and locates the student on the Grades K–6 vertical scale (see Chapter 5 for details).

This report, a portion of which is shown in Figure 8.1, presents performance both numerically and graphically. The bar graph displays the student's overall scale score in relation to:

- the range of scores representing on-grade level performance
- the Mid Grade Level cut established for the academic year based on the student's baseline score (typically the first valid assessment completed during the school year)

As students complete subsequent assessments, the chart also shows changes in scale score since the baseline.

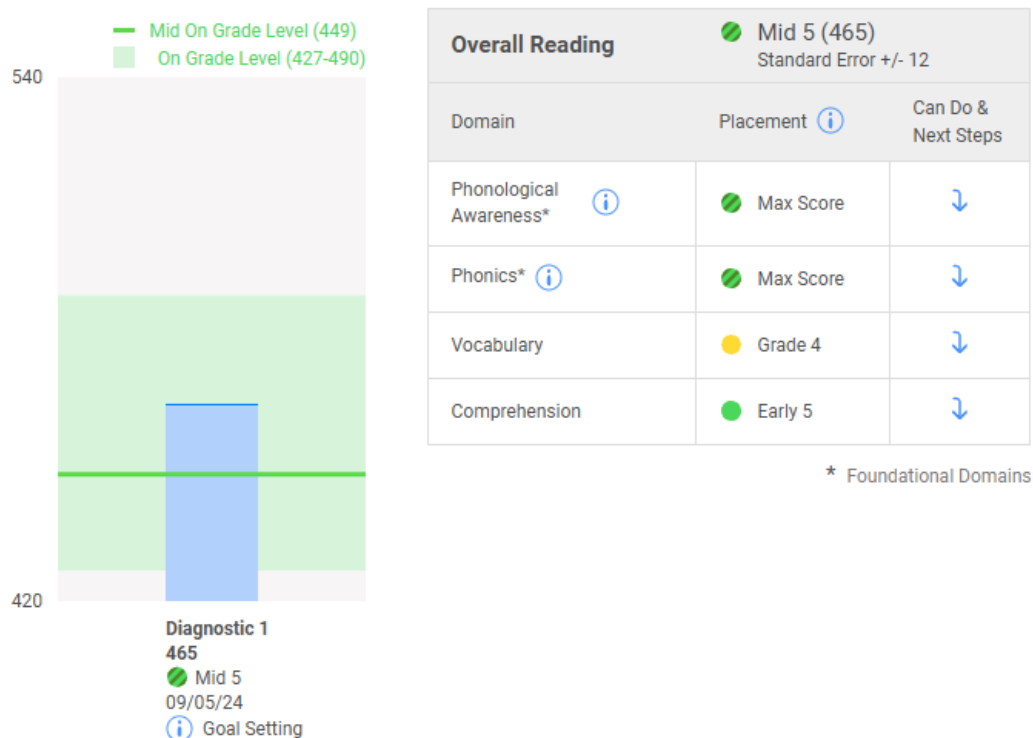


Figure 8.1. i-Ready Inform Results Report, Student Level: Scale Score and Placements

8.3.1.3 Overall Placements, Domain Scale Scores, and Domain Placements

The overall and domain placement levels appear in a table at the top of the report (see Figure 8.1). Next to the overall placement level is the overall scale score—identical to the value in the bar chart—and its standard error of measurement (SEM) (see Chapter 6 for details).

Placement levels are criterion referenced, providing a way to interpret student performance relative to grade-level expectations based on the content standards.

- Students whose overall and domain scores are within their grade-level range receive one of the on-grade level placements.
- Students scoring above or below grade level receive placement designations that reflect the grade-level range in which their score falls. These distinctions are collapsed into Above Grade Level or Below Grade Level or expressed as the number of grade levels above or below.

The placement-level table also includes clickable features that provide online interpretive support.

In the lower portion of the report (Figure 8.2), domain placement levels appear alongside corresponding domain scale scores. An information icon next to each domain title opens a pop-up box explaining the meaning of the placement-level terms and how they inform instruction.

Placement by Domain

Results indicate that Mia F. is decoding accurately but is not comprehending well. The Vocabulary score indicates that word knowledge is not a contributing factor. It is possible that decoding still requires extra effort, which can interfere with reading for meaning. Targeted instruction in key Comprehension strategies will be effective. This information places Mia F. in Instructional Grouping Profile 4.

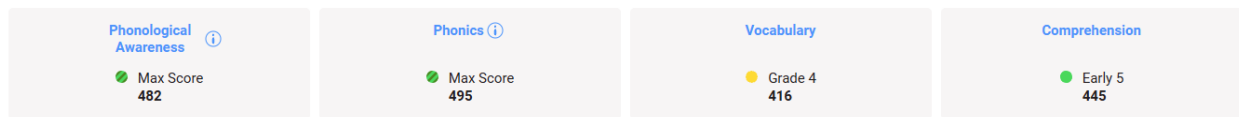


Figure 8.2. *i-Ready Inform* Results Report, Student Level: Domain Scores and Placements

8.3.1.4 Lexile® Measures

i-Ready Inform de lectura also reports El Sistema Lexile® *para Leer* measures for Spanish reading developed by MetaMetrics (Figure 8.3). A Lexile score estimates a student's reading ability based on the complexity of text passages and helps teachers, parents, and students select appropriate reading materials.

Lexile scores are reported as:

- a single value representing the student's reading ability
- a corresponding range to guide text selection

Within the online report, teachers can click on hyperlinks to:

- access a one-page reference guide explaining the meaning and purpose of Spanish Lexile scores
- visit MetaMetrics's resource hub for guidance on interpreting and using Spanish Lexile scores to support instruction

Spanish Lexile® Framework for Reading Measure

Spanish Lexile® Reading Measure:	Spanish Lexile Range:
915L	815L-965L
Understanding Spanish Lexile reading measures 	

The Lexile® Find a Book tool enables you to search for books by grade, interest, and Lexile measure. You can view a book's most challenging words and build a customized reading list. Search for books and see additional Lexile tools now at [Hub.Lexile.com](https://www.hub.lexile.com)

Figure 8.3. *i-Ready Inform* Results Report, Student Level: Norm Percentile and Lexile (Reading Only)

⁶Lexile® measures and the Lexile Framework for Reading are trademarks of MetaMetrics, Inc., and are registered in the United States and abroad. Copyright © 2025 MetaMetrics, Inc. All rights reserved.

8.3.1.5 Can Dos and Next Steps

Can Do statements (“Can Dos”) and **Next Steps** provide actionable, skill-level information to guide instruction (Figure 8.4).

- **Can Dos** describe skills a student is likely able to perform based on their domain proficiency level.
- **Next Steps** identify recommended focus areas for skills that may not yet be mastered. These are based on the finer-grained skills typically—though not exclusively—associated with a grade–domain combination (see Chapter 2).

The difficulty of items associated with each summary claim is used to construct summary claim characteristic curves (similar to test characteristic curves but limited to the items for that summary claim). These curves relate student proficiency to the expected number of items answered correctly for each summary claim and are used to determine summary claim difficulty.

Figure 8.4 shows the academic year 2024–2025 distribution of summary claim difficulties by grade and domain.

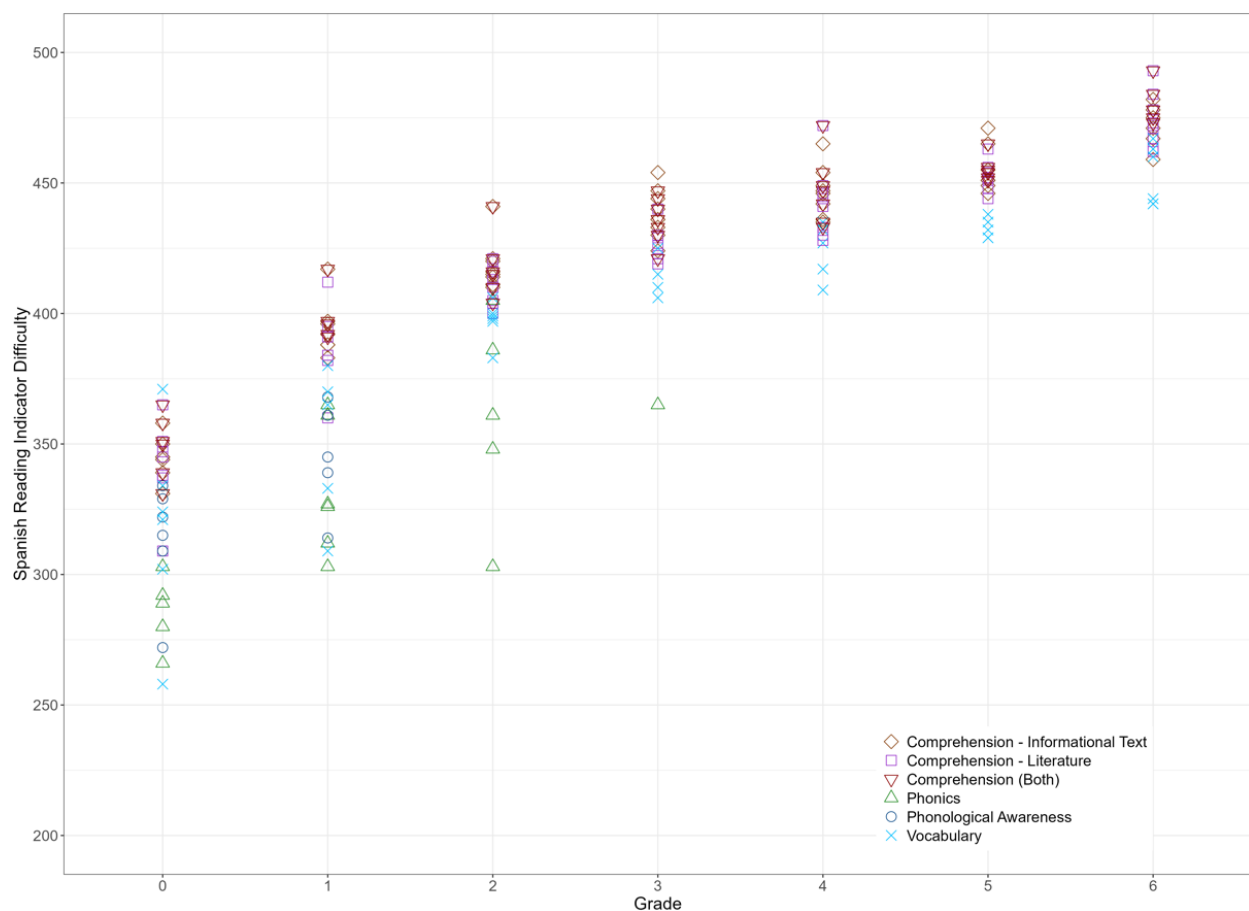


Figure 8.4. Spanish Reading Summary Claim Difficulty Distribution by Grade and Domain

Students' domain placements are compared to the difficulty of each summary claim—provided the claim includes enough items—to determine whether the student is likely to demonstrate the skill represented by that summary claim.

- If a student's domain proficiency is sufficiently high relative to the summary claim difficulty, the skill is considered one the student can likely perform and is labeled as a Can Do on the score report.
- If a student's domain proficiency does not meet this criterion, the skill is labeled as a Next Step.

Skills are prioritized by grade and placement level, so the set of eligible skills for Can Dos and Next Steps varies across students (Figure 8.5).

Developmental Analysis

Both word knowledge and word-learning strategies are addressed in this domain. Mia F. would likely benefit from instruction and practice in more words, particularly vocabulary used in literature and in content area texts such as science, social studies, and math. This student should also receive continued instruction in identifying the context that provides clues to the meaning of academic and domain-specific words and phrases as well as defining the meaning of figurative language, such as similes, metaphors, personification, and idioms.

Can Do ⓘ

Determine the meaning and use of general academic, conversational, and domain-specific words and phrases appropriate to Grade 3.

[Standards](#)

Determine the meaning of words appropriate to Grade 4 based on prefixes, suffixes, and roots.

[Standards](#)

Next Steps & Resources for Instruction ⓘ

- + [Categorize words appropriate to Grade 4.](#)
- + [Identify the context that provides clues to the meaning of academic and domain-specific words and phrases a...](#)
- + [Define the meaning of Grade 4 appropriate figurative language, such as similes, metaphors, personification, a...](#)
- + [Use Grade 4 appropriate words and phrases from general academic and domain-specific contexts, including t...](#)
- + [Identify synonyms and antonyms, appropriate to Grade 4.](#)

Figure 8.5. *i-Ready Inform* Results Report, Student Level: Can Do Statements, Next Steps, and Additional Resources

8.3.1.6 Instructional Guidance and Resources

Instructional guidance based on the student's performance relative to grade-level standards for each domain is accessible by clicking the domain title in the **Placement by Domain** section of the online report. This guidance includes a **Developmental Analysis**, which summarizes the student's performance and identifies areas of instructional readiness and need.

In the **Can Dos & Next Steps** column of the placement-level table (Figure 8.1), a down-arrow hyperlink directs users to the **Developmental Analysis** section. This section provides individualized Can Dos, Next Steps, and links to instructional resources. These features ensure that users notice and easily access instructional resources.

8.3.2. Class Level

The class-level report summarizes the performance of students in a class and is available to teachers and administrators. At the top of the report (Figure 8.6), users see an overview of class performance,

including the number and percentage of students in each placement level based on their overall performance.

Overall Placement

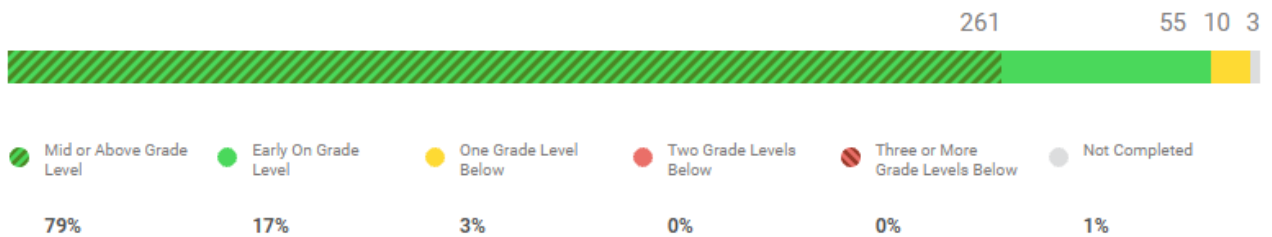


Figure 8.6. *i-Ready Inform* Results Report, Class Level: Overall Placement Level Distribution

Figure 8.7 shows placement-level distribution information displayed by domain.

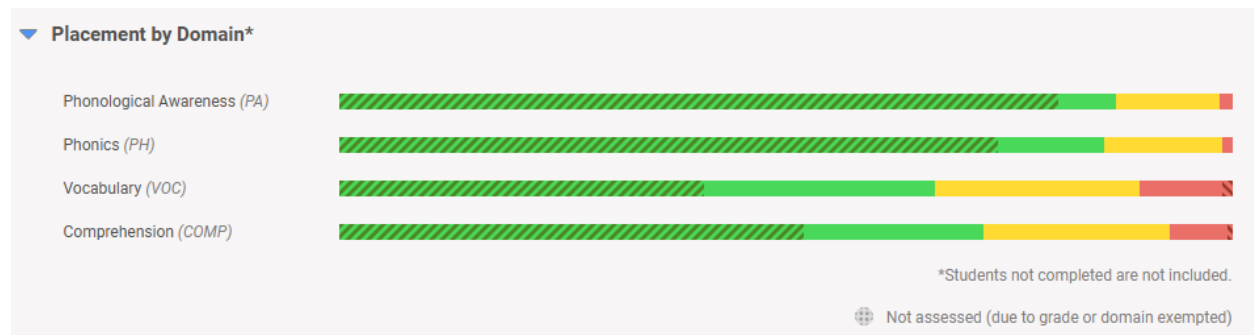


Figure 8.7. *Inform* Results Report, Class Level: Domain Placement-Level Distribution

Finally, a dynamic view (Figure 8.8) shows domain placements for each student. Users can select additional data points to display, such as Lexile® or assessment dates. In 2026-2027, users can also click a yellow **Rush** alert to open a pop-up box explaining the alert.

Student		Placement by Domain				Spanish Lexile® measure & range
Student	Overall Placement & Scale Score	PA	PH	VOC	COMP	Spanish Lexile® measure & range
Patel, Mia F.	Mid 5 (472)	Max Score	Max Score	Grade 3	Grade 6	965L, 865L-1015L
Baker, Danielle	Mid 5 (471)	Max Score	Max Score	Mid 5	Grade 4	960L, 860L-1010L
Simmons, Tristan F.	Mid 5 (471)	Max Score	Max Score	Grade 3	Mid 5	960L, 860L-1010L
Stanton, Geena	Mid 5 (471)	Max Score	Max Score	Mid 5	Early 5	960L, 860L-1010L

Figure 8.8. *i-Ready Inform* Results Report, Class Level: Dynamic and Customizable Table of Student Results

8.4. For Families Report

The For Families report summarizes a student's performance on assessments completed during the school year. It supports teachers in conversations with families by highlighting:

- the student's performance overall and across domains
- skills the student is mastering or areas needing support
- goals for growth
- suggested activities families can do at home to reinforce reading skills

The report is available in English and Spanish. Teachers can print the report to share assessment results with parents. Districts may also enable the report on the student dashboard for efficient distribution. Additional details are provided in the [Understanding Your Student's i-Ready Inform de lectura Results — Reference Sheet](#).

Figure 8.9 and Figure 8.10 illustrate key features of the report.

- Figure 8.9 shows **Overall Spanish Reading Performance**. The summary graph shows results for up to three tests completed during the school year, compared to the grade-level band for context. Each administration displays the scale score, placement level, and growth from the baseline assessment.
- A domain table shows domain performance relative to grade level comparing two assessments (baseline and most recent). This allows users to see where the student excels, has grown, or needs support.
- Figure 8.10 shows the **Skill Progress and More Information** section of the report. This section displays domain-level placements levels along with additional details, such as examples of skills the student has likely mastered and skills to focus on next.

Mia's Overall Spanish Reading Performance

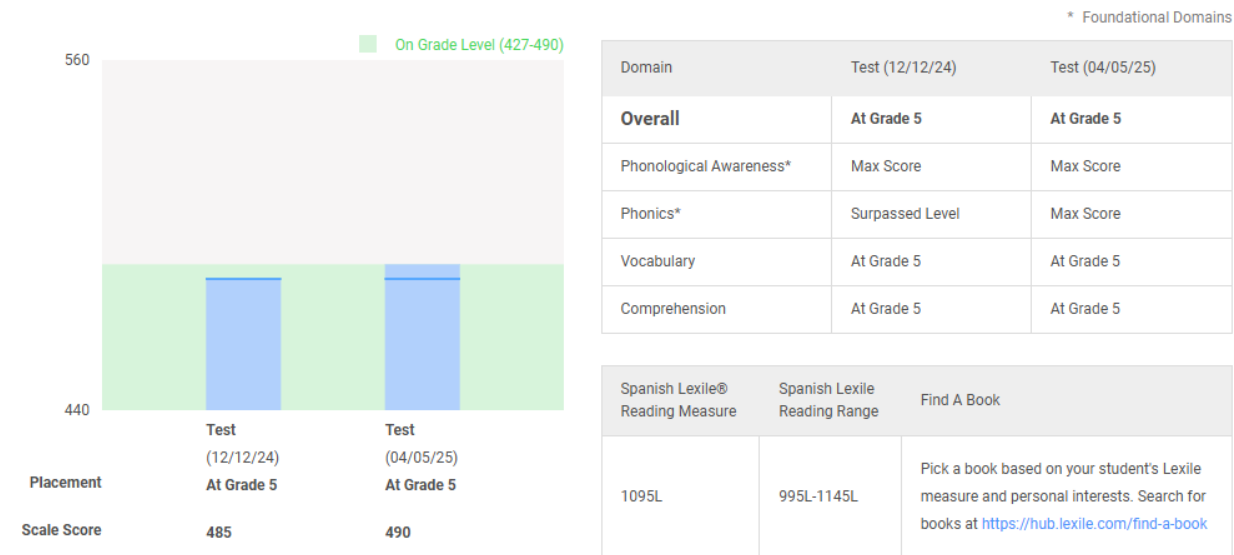


Figure 8.9. Partial View of Page 1 of the For Families Report in Spanish Reading

Skill Progress and More Information

* Foundational Domains ↑ Placement Improved from Baseline

Domain	Test (12/12/24)	Test (04/05/25)	More Information
Phonological Awareness*	Max Score	Max Score	This domain is focused on how students distinguish the sounds (or phonemes) in spoken words. Results indicate that Mia is developing proficiency in the phonological awareness skills that emerging readers need. This student should now focus on the recommendations found in the Phonics domain.
Phonics*	Surpassed Level	Max Score	This domain focuses on how accurately students decode written words. Results indicate that Mia is achieving accuracy in recognizing spelling patterns, but this student should continue to practice decoding and reviewing multisyllabic words to be able to decode fluently.
Vocabulary	At Grade 5	At Grade 5	Both word knowledge and word-learning strategies are addressed in this domain. Mia is ready for instruction and practice with determining the meaning of words based on prefixes, suffixes, and roots. In addition, the student should continue to expand and deepen knowledge of vocabulary used in grade-level appropriate literature as well as in science, social studies, and math texts.
Comprehension ↑	At Grade 5	At Grade 5	This domain addresses Mia's understanding of literary and informational text. Results indicate that Mia is ready for instruction in Grade 5 comprehension skills and strategies such as drawing conclusions or making inferences, analyzing how chapters in a story fit together to create the overall structure, and determining an author's purpose or perspective. Support the student in reading a variety of genres including realistic and historical fiction, fables, tall tales, and myths, as well as biographies, reference sources, and opinion essays. Provide numerous opportunities for Mia to summarize text.

Figure 8.10. Partial View of Page 2 of the For Families Report

8.5. Additional Interpretive Guidance

Curriculum Associates provides educators with clear, actionable information about *i-Ready Inform de lectura*—how it works and how to interpret results. In addition to the guidance embedded in each report, users and stakeholders have access to:

- an extensive library of [Frequently Asked Questions](#)
- *i-Ready Success Central* resources to support accurate interpretations of assessment results

Each *i-Ready* user is supported by an *i-Ready Partners* team, which includes professional learning specialists and partner success managers. These experts focus on successful implementation, and depending on account size, may offer real-time reviews of achievement data after each testing window (beginning, middle, and end of year). These reviews provide detailed analyses of student performance to inform data-driven practices and action planning.

In alignment with the *Standards for Educational and Psychological Testing* (AERA et al., 2014), particularly Standard 12.15, Curriculum Associates offers professional learning, one-on-one support, and training courses through the Online Educator Learning platform to promote valid interpretations and sound use of data.

i-Ready provides extensive instructional resources within the platform, directly accessible from student and class results, along with guidance for using these resources effectively. Support materials are tailored for intended audiences, consider the full range of grade levels, and help educators use *i-Ready Inform de lectura* results to differentiate instruction for students across the performance spectrum.

Chapter 9. Validity

9.1. Chapter Summary

This chapter presents a validity argument for *i-Ready Inform™* de lectura, grounded in the framework from the *Standards for Educational and Psychological Testing* (AERA et al., 2014) and the evidence that supports the validity argument. Validity is defined as the degree to which evidence and theory support the intended interpretations and uses of test scores.

Five sources of validity evidence are summarized: test content, response process, internal structure, relationships with other variables, and consequences of testing. The chapter synthesizes information from earlier chapters of this technical manual with new findings introduced here. Table 9.1 maps chapters to claims about score interpretation and intended use.

Then, throughout the chapter, we summarize this information and organize it into the five sources of validity evidence described by the Standards and show how it supports the validity argument for *i-Ready Inform* de lectura.

9.2. Introduction

As outlined in Chapter 1, validation is a process of developing and substantiating a validity argument (Kane, 2006, 2013). Validation focuses on gathering evidence to support the intended uses and interpretations of scores, while validity is the degree to which evidence and theory support the interpretations of test scores for intended uses (AERA et al., 2014). This chapter integrates evidence presented in previous chapters with additional analyses to offer a comprehensive review of the validity argument for *i-Ready Inform* de lectura.

9.2.1. Five Sources of Evidence

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) identify five categories for organizing validity evidence to develop a strong validity argument. This chapter is organized around these categories, which are briefly described below.

Evidence Based on Test Content. This category includes evidence that the assessment measures the breadth of the intended construct through its content. Examples of supporting evidence include test specifications, item writing protocols, content organization schemata, and standardized administration procedures.

Evidence Based on Response Process. This evidence supports the inferential link between the assessment activity (i.e., letter, object, word) and the construct being measured. The strength of this inference varies depending on the nature of the assessment activity and the construct. In cases where

only response correctness is of interest, no inferential link is required (AERA et al., 2014, p. 16). Common sources of evidence include item development and item selection.

Evidence Based on Internal Structure. This category demonstrates that the assessment’s statistical properties align with the construct being measured and are of sufficient quality to support the interpretations and uses of the scores. Key indicators of evidence include reliability analyses.

Evidence Based on Relationships with Other Variables. This category of evidence includes two subtypes:

- **Convergent and discriminant (also known as divergent) relationships** with other measures and test-criterion relationships (AERA et al., 2014). Such relationships show that the instrument correlates strongly with other instruments of similar constructs (convergent) and does not correlate strongly with measures of dissimilar constructs (discriminant).
- **Test-criterion relationships** demonstrate a relationship between the instrument and an external criterion. Examples of evidence include linking studies.

Evidence Based on Consequences of Testing. This category addresses both the intended and unintended consequences of testing, evaluating whether those consequences are justifiable and align with validity claims. *i-Ready Inform de lectura* is a key component of the Theory of Action (See Chapter 1), whose outcomes drive reporting and guidance on the use of instructional resources to address skill development. Evidence in this category may depend on how faithfully other components of the Theory of Action are implemented.

9.2.2. Locating the Evidence

To organize and evaluate the validity evidence within this manual, we examine how each chapter provides evidence to support the validity claims. **Error! Reference source not found.** and Table 9.2 identify the source of evidence presented by chapter and claim. Table 9.1 focuses on score interpretation, while Table 9.2 addresses intended uses of scores. The remaining sections of this chapter provide a brief narrative of the evidence, organized by type.

Table 9.1. Sources of Validity Evidence by Chapter and Claim: Score Interpretation				
Chapter and Title		Claims of <i>i-Ready Inform de lectura</i>		
		Interpretation of Scores		
		Measures Domains in Spanish Reading	Reports a Variety of Estimates of Accurate Student Proficiency	Contextualizes Student Proficiency
1	Introduction	T		

Table 9.1. Sources of Validity Evidence by Chapter and Claim: Score Interpretation					
Chapter and Title		Claims of <i>i-Ready Inform de lectura</i>			
		Interpretation of Scores			
		Measures Domains in Spanish Reading	Reports a Variety of Estimates of Accurate Student Proficiency	Contextualizes Student Proficiency	
2	Test Design and Development	T R		T I	
3	Test Administration	T	T		
4	CAT and Test Flows	T R	T		
5	Calibration and Scaling	I	I	I	
6	Reliability				
7	Setting Standards	T C	T C	T I C	
8	Score Reports		C	C E	
9	Validity	E		E	

Note: T = Test Content; I = Internal Structure; R = Response Process; E = Relationships with External Variables; C = Consequences

Table 9.2. Sources of Validity Evidence by Chapter and Claim: Intended Score Uses					
Chapter and Title		Claims of <i>i-Ready Inform de lectura</i>			
		Intended Uses of Scores			
		Identify and Monitor Students	Provide Feedback on What Students Can Likely Do and Should Work on Next	Group Students to Provide Differentiated Instruction	Route Students into the Appropriate Personalized Instruction Module
1	Introduction			C	C
2	Test Design and Development		I		
3	Test Administration				
4	CAT and Test Flows				
5	Calibration and Scaling			I	I
6	Reliability			I	I
7	Setting Standards	T C	I	T C	TC

Table 9.2. Sources of Validity Evidence by Chapter and Claim: Intended Score Uses					
Chapter and Title		Claims of <i>i-Ready Inform de lectura</i>			
		Intended Uses of Scores			
		Identify and Monitor Students	Provide Feedback on What Students Can Likely Do and Should Work on Next	Group Students to Provide Differentiated Instruction	Route Students into the Appropriate Personalized Instruction Module
8	Score Reports	C	C	C	C
9	Validity	E	E		

Note: T = Test Content; I = Internal Structure; R = Response Process; E = Relationships with External Variables; C = Consequences

9.3. Evidence Based on Test Content

Evidence related to test content concerns the relationship between the assessment’s content and the construct it is intended to measure. Test content refers to the subject matter, wording, and the format of items as well as their administration and scoring. Test specifications describe the test content in detail and provide evidence to support the intended interpretation of the scores. Test administration documentation also supports the intended interpretation of the scores.

9.3.1. Assessment Design and Development

Chapters 1 and 2 describe the foundations of *i-Ready Inform de lectura* and our principled approach to assessment design and development, including:

- a description of the target population
- how the test construct (Spanish reading) is operationalized through construct mapping, summary claims, and placement level descriptors (PLDs)
- item formats used and how items are specified, written, reviewed, and aligned to content and skills
- how items are field tested (see Chapter 5) to evaluate psychometric item characteristics prior to inclusion in operational item pools
- What types of items are administered
- how content is sampled (i.e., test flows, see Chapter 4) and the cognitive processes elicited by items

- how we ensure students have full access to the assessment content, including a discussion of how accessibility supports and accommodations are incorporated into the assessment design

The outcomes of the content development process provide validity evidence for *i-Ready Inform de lectura*, ensuring that the alignment between the construct and test content is clearly articulated and supported by a sound rationale and accompanying evidence. The evidence presented in Chapter 2 establishes the content-based rationale for using *i-Ready Inform de lectura* to assess achievement in Spanish reading, supporting intended interpretations and uses. Although not a feature of assessment design or content, administration conditions are critical to ensuring score comparability across students and over time. Student responses form the basis of score reports and psychometric components of *i-Ready Inform de lectura* (i.e., raw scores, scaling, equating, and score scales), which are documented throughout this technical manual.

9.3.2. Task Administration and Accommodations

Although not a feature of assessment design or content, administration conditions are critical to ensuring score comparability across students and over time. Student responses form the basis of the score reports and psychometric components of the assessment, which are documented throughout this manual.

Chapter 3 provides evidence of test administration conditions and supports available to students. It provides clear guidance to ensure consistent practices that accurately reflect student knowledge and proficiency. The chapter details the testing windows (fall, winter, spring), recommended testing times, and procedures for test proctors and administrators to prepare for, conduct, monitor, and complete an administration of *i-Ready Inform de lectura*. It also describes accessibility features and allowable accommodations for students with disabilities and multilingual learners, helping to minimize construct-irrelevant barriers to student performance.

9.3.3. The Adaptive Algorithm and Test Flows

Chapter 4 explains how content specifications are embedded within the item selection algorithm. The algorithm applies adaptive assessment methodology to optimize score precision by selecting items closely matched to students' estimated proficiency levels. This approach enhances reliability and reduces the standard error of measurement (SEM) compared to a linear (non-adaptive) test of the same length, assuming all other factors remain constant. In this way, the algorithm supports appropriate interpretations and effective uses of the scores.

Test flows are integrated into the algorithm to ensure that content presented to students is appropriate for their grade level and performance on earlier items. The chapter also summarizes simulation research conducted by Curriculum Associates to refine processes and improve the effectiveness of *i-Ready Inform de lectura*. This research uses a simulator to evaluate the efficiency of decisions related to item selection, proficiency estimation, and item development priorities, supporting ongoing item bank maintenance.

9.3.4. Criterion-Referenced Placement Levels

Placement levels form the basis for criterion-referenced score interpretations. Chapter 2 describes the placement levels, and Chapter 7 outlines the standard-setting methodology used to determine the cut scores between them. Performance Level Descriptors (PLDs) provide a detailed framework for mapping content to items. Two key intended uses of scores—instructional groupings and Personalized Instruction—are based on placement levels and guide the delivery of appropriate instructional content to students.

9.4. Evidence Based on Response Process

Evidence related to response processes focuses on the cognitive processes students use while completing the assessment. This evidence may include theoretical and empirical analyses of the response processes to evaluate the alignment between the intended construct and the cognitive processes elicited by test items, particularly where a substantial inferential link is required.

9.4.1. Measures of Cognitive Complexity and Difficulty

Cognitive complexity, or cognitive demand, is a key specification when developing items for *i-Ready Inform de lectura*. Each assessment domain specifies the types and levels of cognitive complexity necessary to assess the domain. These specifications then guide item development, as described in Chapter 2. For Spanish reading, this includes Webb’s Depth of Knowledge (DOK) framework (Webb, 1997, 1999). The chapter explains how these measures are defined and how methods are applied to ensure item development aligns with intended cognitive complexity.

Chapter 4 further details how cognitive complexity informs item selection. Test flows specify how the adaptive algorithm ensures that all students encounter an appropriate range of cognitive demand. The computer adaptive test (CAT) simulator evaluates changes to the item bank, item selection methods, and test flows to maintain optimal item selection and support ongoing item bank maintenance.

9.5. Evidence Based on Internal Structure

The internal structure of an assessment refers to the interrelationships among its items and components. A consistent internal structure supports the idea that test items measure the same construct or closely related constructs.

9.5.1. Measurement Characteristics

Chapter 5 presents the psychometric theory and supporting research underlying *i-Ready Inform de lectura* score scales and proficiency estimation. Evidence includes:

- application of Classical Test Theory (CTT) to monitor item psychometric properties and flag items with unexpected or unfavorable characteristics
- use of Item Response Theory (IRT) for calibration, equating, and vertical scaling
- field testing to refresh the item bank and maintain a single bank across administrations
- calibration studies to establish the vertical scale

These practices demonstrate adherence to best practice approaches to develop the *i-Ready Inform* delectura score scale. The scale provides a meaningful metric rooted in criterion-referenced placement levels (Chapter 7) and aligned with the PLDs (Chapter 2). Together, these elements support valid interpretations of student knowledge across content domains and ensure that score uses are consistent with the overall validity argument.

9.5.2. Reliability

Marginal reliability estimates offer insights into score precision and are computed for the overall assessment. These estimates are also calculated for select student groups. Reliability analyses indicate the cohesiveness and integrity of the test's internal structure, which—combined with evidence of strong test design and development—supports the conclusion that test items consistently measure the intended construct.

Chapter 6 reports that overall marginal reliability estimates are .90 or higher for all students tested during the fall 2025 testing window (August to mid-October). For select student groups, marginal reliability estimates are .88 or higher. Corresponding SEMs are also presented. Reliability values at or above .90 indicate strong measurement precision, providing high confidence that placement decisions and score interpretations are consistent and dependable across administrations.

9.5.3. Classification Accuracy and Consistency

Classification accuracy and consistency analyses support the validity of placement levels by demonstrating sufficient reliability of classification. Because many intended uses of scores are based on placement levels, evidence that students are accurately and consistently classified provides validity evidence for these use cases.

Results in Chapter 6 show strong classification accuracy for each individual cut score and moderate to strong classification when all five typical placement levels are considered together. Classification accuracy for single cut scores is $\geq .90$ and $\geq .75$ for all cuts combined. Classification consistency for all five typical cuts is $\geq .67$, which is slightly lower than classification accuracy, as expected. These findings indicate that placement decisions based on cut scores are highly dependable, providing strong confidence that students are classified accurately and consistently for instructional purposes.

9.6. Evidence Based on Relationships to Other Variables

This section summarizes the validity evidence for *i-Ready Inform de lectura* scores based on their relationships to scores from other instruments, focusing on how scores align with both similar and dissimilar external measures. It presents analyses of *convergent* validity (relationship with assessments that measure similar content) and *discriminant* validity (relationship with assessments that measure dissimilar content). Together, these analyses support the intended use of *i-Ready Inform de lectura* by demonstrating how well it aligns with established assessments.

9.6.1. Linking Studies

Evidence based on relationships with other variables comes from linking studies conducted by Curriculum Associates in collaboration with MetaMetrics. These studies link *i-Ready Inform de lectura* performance to the Spanish Lexile Framework for Grades K–6 and provide convergent validity evidence for measures administered close in time. Spanish Lexile scores are available on student reports, shown in Chapter 8.

For the linking studies, students in Grades 1, 3, and 5 completed a standalone MetaMetrics Lexile form in addition to *i-Ready Inform de lectura* items. Because *i-Ready Inform de lectura* uses a vertical scale, a Spanish Lexile link could be established for Grades K–6. The overall correlation between the two assessments across grades was .90. This high correlation reinforces the test’s external validity and provides strong evidence that the *i-Ready Inform de lectura* scale aligns closely with the Spanish Lexile scale, supporting accurate interpretation.

9.6.2. Correlation Studies

Correlation studies examine the relationships between *i-Ready Inform de lectura* scores and scores from other assessments. During the fall 2025 testing window, many students completed both *i-Ready Inform de lectura* and *i-Ready Inform for Reading*. The correlation between the two assessments across grades was .65, which is expected because the assessments measure different constructs. These results provide evidence of discriminant validity. Moderate correlations confirm that while the two assessments share some underlying skills, they remain distinct measures, supporting appropriate interpretation and use of each score for its intended purpose.

9.7. Evidence Based on Test Consequences

Evidence based on test consequences refers to the evaluation of how the use and interpretation of assessment results impact students, instructional practices, and educational outcomes. In psychometric research, this type of evidence is essential for validating not only the technical quality of a test but also its real-world effects—such as whether students are appropriately identified for intervention, whether resources are allocated effectively, and whether the assessment supports equitable educational opportunities (Lane, 2014; AERA et al., 2014).

Consequential validity evidence evaluates the consequences of using test scores. This manual provides detailed information on test content and technical quality, supported by extensive research to justify intended interpretations and uses of scores. Much of this evidence is reflected in score reports (see Chapter 8). In alignment with the intended uses outlined in Chapter 1, Chapter 7 describes the standard setting processes that determine cut scores for placement levels, which directly influence student placement and the instructional resources provided to teachers.

9.8. Discussion

This chapter concludes the validity discussion that began in Chapter 1 with our Theory of Action. In summary, the validity evidence presented in this technical manual supports the assessment claims outlined in Chapter 1. We organized this evidence using the five sources of validity evidence described by AERA *et al.* (2014), as depicted in Figure 1.2.

Curriculum Associates continually collects evidence of reliability to support validity claims. In addition to conducting planned reliability and validity studies, we analyze naturally occurring data from students to examine how the tasks function in classroom settings, especially as assessments are updated or revised, as new evidence to support the validity argument is warranted.

As *i-Ready Inform* de lectura continues to be administered, additional opportunities will arise to collect further validity evidence. Future studies may include:

- Evidence Based on Test Content
 - Alignment studies
- Evidence Based on Response Process
 - Cognitive laboratory studies
 - Monitoring response process times
- Evidence Based on Internal Structure
 - Differential item functioning
- Evidence Based on Relationships to Other Variables
 - State summative assessment linking studies
- Evidence Based on Test Consequences
 - Growth measures
 - Efficacy studies
 - Development of user norms

References

- American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. AERA.
- Brice, A. & Mix, D. (2021, May). Practical application of principled assessment design in building content expertise while developing and refining PLDs through embedded standard setting. In Davis, L. (Chair) Creating coherence: Integrating principled assessment design, PLDs, and standard setting. Session conducted at the annual meeting of the National Council on Measurement in Education, Virtual.
- Briggs, D. C. & Dadey, N. (2015). Making sense of common test items that do not get easier over time: Implications for vertical scaling designs. *Educational Assessment*, 20(1), 1-22.
- Cappaert, K., Morrison, K., & Su, Y. (2022). Relaxed CAT: An extended b-matching method for content relaxed item selection. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- CAST. (2024). *Universal design for learning guidelines version 3.0*. <https://udlguidelines.cast.org>.
- Egan, K. L., Schneider, M. C., & Ferrara, S. (2012). Performance-level descriptors: History, practice, and a proposed framework. In G. Cizek (Ed.), *Setting performance standards* (2nd ed., pp 79–106). Routledge.
- Huff, K., Nichols, P., & Schneider, C. (2025). Designing and developing educational assessments. In L. L. Cook & M. J. Pitoniak (Eds.), *Educational measurement* (5th ed., pp. 441–512). National Council on Measurement in Education.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). American Council on Education/Praeger.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kolen, M. J. & Brennan, R. L. (2010). Test equating, scaling, and linking: Methods and *practices* (2nd ed.). Springer.
- Laitusis, C. C. & Karvonen M. (2024, April). *A case for reimagining universal design in assessment systems*. Paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA.
- Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, 26(1), 127–135. <https://psycnet.apa.org/record/2014-03588-019>

- Lewis, D. & Cook, R. (2020). Embedded standard setting: Aligning standard setting methodology with contemporary assessment design principles. *Educational Measurement: Issues and Practices*, 39(1), 8–21.
- Lewis, D., Lee, S., & Choi, S. (2021). *An investigation of two embedded standard setting cut score estimation algorithms: ESS-Count and ESS-Weight*. In D. Lewis (Chair), Embedded standard setting: Research and advances. Session conducted at the annual meeting of the National Council on Measurement in Education, Virtual.
- Linacre, J. M. (2023). Winsteps (Version 4.4.0) [Computer software] Winsteps.com.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. <https://doi.org/10.17226/10019>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danmarks Paedagogiske Institut: Copenhagen.
- Rudner, L. M. (2000). Computing the expected proportions of misclassified examinees. *Practical Assessment, Research, and Evaluation*, 7(14). <https://doi.org/10.7275/an9m-2035>
- Rudner, L. M. (2005). Expected classification accuracy. *Practical Assessment, Research, and Evaluation*, 10(1), 13.
- Ryan, R. M., & Deci, E. L. (2018). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press. <https://doi.org/10.1521/978.14625/28806>
- Sireci, S.G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237–247.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6). Madison: University of Wisconsin, Wisconsin Center for Education Research.
- Webb, N. L. (1999). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states: A study of the state collaborative on assessment & student standards (SCASS)*, Technical Issues in Large-Scale Assessment (TILSA). Council of Chief State School Officers.
- Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7–25.
- Worldwide Web Consortium. (2025). *Web content accessibility guidelines (WCAG)*. W3C: Web accessibility initiative (WAI). <https://www.w3.org/WAI/standards-guidelines/wcag>

Wyse, A. E., & Hao, S. (2012). An evaluation of item response theory classification accuracy and consistency indices. *Applied Psychological Measurement*, 36(7), 602-624.
<https://journals.sagepub.com/doi/pdf/10.1177/0146621612451522>