The Literably Screener
Technical Manual

Literably

Table of Contents

Table of Contents **2**
Technical Report 4
Test Development 4
Intended Uses 4
Content Requirements 4
Content Specifications 5
Overview 5
Rapid Automatized Naming (RAN) 6
Phonological Awareness 8
Phonics 9
Spelling 12
Oral Reading Fluency 15
Vocabulary 16
Reading Comprehension 19
Scoring Method 20
Scoring Method 20
Special Scoring Considerations 25
Reliability 27
Subtest Reliabilities 27
Composite Score Reliability 28
Delayed Alternate-Form Reliability 29
Standard Error of Measurement 30
Validity 30
Content Validity 30
Construct Validity 40
Criterion Validity 44
Classification Accuracy 47
Definition of Risk 47
Cut Score Methodology 47
Composite Cut Scores and Classification Accuracy Estimates 47
Subtest Cut Scores and Classification Accuracy Estimates 49
Bias Analyses 52
Differential Item Functioning 52
Composite Validity by Subgroup 53
Item Difficulty 55
Interpretation of Scores 55
References 57
Test Administration 61
Overview 61

Administration Timeline and Mode....61
Technology Requirements....62
Administration Protocol....63
Testing Irregularities....68
Test Administrators....69
Trainings and Resources....69
Accessibility Features....70
Description of Test Security Measures....74
Confidentiality of Student Data....74
Assessment Integrity....74
Sample Test Forms and Score Reports....76
Sample Test Items....76
Sample Score Reports....81

Technical Report

Literably, a subsidiary of Heggerty, is a for-profit Delaware C-Corporation founded in 2013 to provide reading assessment services to schools. Literably offers literacy assessment products and services designed to screen students for reading difficulties and inform appropriate instruction. This report was compiled by staff members at Literably, with advice from leading literacy experts and research support from WestEd.

Test Development

The Literably Screener is a universal screening instrument that measures students' acquisition of literacy skills from kindergarten through fifth grade. It is designed to pinpoint students who may be at risk of reading difficulties and identify areas to target for instructional support. The Screener is computer-administered and accessible via desktop, laptop, and tablet devices. All items are administered and scored by Literably. The screener is easy to administer, reliable and valid, and yields actionable data to drive instructional next steps.

Literably's development process adheres to best practice testing standards (AERA et al., 2014), including:

- Specifying intended uses
- Defining content requirements
- Creating and selecting items according to specifications
- Iterating based on internal and external review and feedback

Intended Uses

The Literably Screener is intended for use up to three times per year from the beginning of kindergarten through the end of fifth grade. Literably is intended for two purposes:

- To identify students who may be at risk of severe reading difficulties, including dyslexia
- To identify students who are not on track to achieve grade-level literacy goals

Content Requirements

By design, the Literably Screener measures constructs that are well-recognized, research-based indicators of foundational early literacy skills. In order to develop the content requirements for the Literably Screener, Literably researchers performed a comprehensive literature review to identify the literacy measures that are most appropriate when screening for reading difficulties and achievement of grade-level literacy goals. This literature review included:

- Relevant research literature on appropriate screening measures (e.g. National Research Council, 1998; National Institute of Child Health and Human Development, 2000; Rand, 2002; Rayner, Foorman, Perfetti, Pesestsky & Seidenberg, 2001).
- State standards
- State literacy screening requirements
- Technical manuals from existing well-regarded screening assessments

Based on this review, Literably researchers identified a preliminary list of subtests for further review. Literably then asked external experts (i.e., professors and researchers with a focus on early literacy assessment) to provide feedback on the preliminary list of subtests. Based on this feedback, Literably arrived at a consensus list of 7 subtests deemed to capture the most critical

constructs predictive of later reading success: Rapid Automatized Naming (RAN), Phonological Awareness, Phonics, Spelling, Oral Reading Fluency, Vocabulary, and Comprehension.

For each grade K-5, the Literably Screener includes a subset of the 7 subtests that literacy experts deem most important when screening at that grade. Within each grade, the screener includes the same subtests at the beginning, middle, and end of the year to allow for ease of administration and consistent data collection. Taken together, the Literably Screener subtests contribute to an evidence-based and grade-appropriate screening instrument that appropriately measures the key constructs that contribute to reading development.

Table 1 below presents the literacy constructs measured by each of the Literably Screener subtests and the grade levels at which these subtests are administered.

***Table 1**. Literably Subtests, Literacy Constructs, and Grade Levels*

Literably Screener Subtest	**Literacy Construct(s)**	**Grade Levels**
Rapid Automatized Naming (RAN)	Rapid Automatized Naming	K-1
Phonological Awareness	Phonological and Phonemic Awareness	K-1
Phonics	Letter Naming, Letter-Sound Correspondence, Real Word Reading, Nonword Reading	K-3
Spelling	Encoding	1-5
Oral Reading Fluency	Oral Text Reading Accuracy and Rate	1-5
Vocabulary	Vocabulary (Receptive)	K-5
Comprehension	Reading Comprehension	2-5

Content Specifications

Overview

The Literably Screener employs tasks that authentically and directly measure the key early literacy constructs that contribute to reading success. For skills that involve the production of oral language (e.g., phonological awareness, phonics, oral reading fluency, RAN), Literably collects students' oral responses for scoring. For skills such as spelling, receptive vocabulary, and reading comprehension, Literably collects students' inputs in the form of multiple choice or typed responses. *Table 2* below summarizes the Literably Screener subtests, constructs measured, and student response modes.

***Table 2.** Literably Subtests and Student Response Modes*

Subtest	**Construct(s)**	**Student Response Mode**

Rapid Automatized Naming (RAN)	Rapid Automatized Naming	**Verbal Response**: Students name 2 arrays of 50 objects each out loud.
Phonological Awareness	Phonological and Phonemic Awareness	**Verbal Response**: Students are presented orally with words and sounds. Students isolate, delete, and substitute sounds in words. Students segment words into parts and blend sounds into words.
Phonics	Letter Naming, Letter-Sound Correspondence, Real Word Reading, Nonword Reading	**Verbal Response**: Students identify the names and sounds of letters and read individual real and nonwords aloud.
Spelling	Encoding	**Typed Response**: Students spell words one at a time using the keyboard on the screen or the keyboard on their device.
Oral Reading Fluency	Oral Text Reading Accuracy and Rate	**Verbal Response**: Students read grade-level passages aloud and select the best answer from 4 answer choices for each question.
Vocabulary	Vocabulary (Receptive)	**Multiple-Choice Selection**: Students select the best answer from 4 answer choices for each question.
Comprehension	Reading Comprehension	**Multiple-Choice Selection**: After silently reading a passage, students select the best answer from 4 answer choices for each question.

For each identified subtest, Literably content experts drafted instructions and item templates via an iterative process including multiple expert reviewers. Once the item templates were established, Literably content experts wrote items according to the item templates, with each item reviewed by multiple expert reviewers. The sections below describe the following as they pertain to each of the 7 subtests on the Literably Screener:

- Theoretical framework
- Item development process
- Test blueprint and content specifications
- Tasks and response modes

Rapid Automatized Naming (RAN)

Since Denckla's first RAN study in 1972, hundreds of studies have supported the use of RAN as a screener for reading difficulties and a predictor of reading growth (e.g., Lervåg & Hulme, 2009; Araújo et al., 2015; Al Otaiba & Fuchs, 2002; Nelson et al., 2003). A recent meta-analysis showed that kindergarten RAN assessment predicts grade-school reading performance, even controlling for differences in phonological awareness (McWeeny et al., 2022), and this predictive relationship persists from kindergarten through adulthood (Hjetland et al., 2017). Research

strongly supports screening for RAN in Grades K-1 (Georgiou, et al., 2011), and the Literably Screener includes a RAN subtest at Grades K-1 to provide more robust screening data in these early grades.

Literably developed Literably RAN–including its format and directions–based on an extensive review of the RAN literature. Literably RAN was field-tested by experienced educators in California, Washington, Massachusetts, Michigan, and Pennsylvania, and enhancements were made in consultation with these educators.

According to McWeeny et al. (2022), a “true RAN task” is one in which items to be named are highly familiar or automatized for students, and where items are arranged in an array and named left-to-right, row-by-row. The Literably RAN subtest conforms to these characteristics by measuring students’ ability to quickly name aloud items from arrays composed of repeating, randomly-ordered groups of five distinct familiar items. Each task array contains five rows of ten items.

Literably RAN includes two typical RAN tasks: letters and numbers. A meta-analysis of RAN results showed that RAN tasks using letters and numbers have the highest correlation with overall reading skill, compared with tasks that use colors or pictures (Araujo, Reis, Magnus Petersson, & Faisca, 2015).

During the Literably RAN design process, Literably test developers carefully selected numbers (4, 1, 7, 8, 2) and letters (T, X, A, J, U) that students would be unlikely to confuse with each other, based on educator input. The stimulus items and example arrays for each RAN task are displayed in *Table 3*.

Table 3: *RAN Tasks, Stimulus Items and Sample Arrays*

Task	**Stimulus Items**	**Sample Array**
Numbers	7, 1, 8, 2, 4	7 1 8 2 4 8 7 4 2 1 8 1 2 7 4 1 7 4 2 8 4 8 7 1 2 7 8 1 2 4 2 4 1 7 8 2 7 8 4 1 2 8 7 4 1 8 1 4 2 7
Letters	A, T, U, X, J	T X J U A T U A J X T X J A U T J X A U X A J U T A X T U J X T J A U T J X U A J T U A X J A U X T

The RAN subtest is a computer-administered task. Students provide their responses orally, and the responses are recorded by Literably in the form of audio files and scored.

To obtain a valid score from an assessment such as RAN, it is important to administer the RAN task according to standardized instructions. Because Literably RAN is a computer-administered assessment, it delivers the same instructions to students each time and thus minimizes variation in the assessment experience.

Phonological Awareness

The Literably Screener Phonological Awareness subtest assesses phonological and phonemic awareness at grades K and 1. Students with strong phonological and phonemic awareness are able to discern and manipulate phonemes and larger units of speech, which contributes to reading readiness by enabling the effective learning and recall of grapheme-phoneme correspondences. A strong body of research shows that phonological awareness is closely related to the acquisition of reading skills (e.g., Ehri, et al., 2001) and that measurements of phonological awareness provide a robust predictor of reading achievement (e.g., Lonigan, Burgess, and Anthony, 2000).

The Phonological Awareness subtest is a direct measure of students' ability to identify and manipulate sounds in spoken words in grades K-1. *Table 4* below lists the number and types of items on the Phonological Awareness subtest at each grade. The content blueprint within each grade is the same across seasons.

Table 4. *Phonological Awareness Test Blueprint*

Grade	**Item Type (Number of Items)**
Kindergarten (20 total items)	Isolating initial phoneme (2) Isolating final phoneme (1) Isolating medial phoneme (1) Blending 3 phonemes (2) Blending 4 phonemes (2) Segmenting 3 phonemes (2) Segmenting 4 phonemes (2) Deleting part of compound words (1) Deleting initial phonemes (3) Substituting initial phonemes (1) Substituting final phonemes (2) Substituting medial phonemes (1)
First Grade (20 total items)	Isolating initial phoneme (2) Isolating final phoneme (1) Isolating medial phoneme (1) Blending 3 phonemes (2) Blending 4 phonemes (2) Segmenting 3 phonemes (1) Segmenting 4 phonemes (2) Segmenting 5 phonemes (1) Deleting initial phonemes (3) Deleting final phoneme of initial blend (1)

	Substituting initial phonemes (1) Substituting final phonemes (2) Substituting medial phonemes (1)

All items in the Phonological Awareness item bank were reviewed by multiple content experts for age-appropriateness, and item statistics were gathered to evaluate their performance. To ensure age-appropriate prompts, all words that appear in the Phonological Awareness item prompts were analyzed for their age of acquisition (Kuperman et al., 2012; Dale & O'Rourke, 1981).

The items on the Phonological Awareness subtest progress from earlier-developing phonological awareness skills to those that develop later, according to research on the general continuum of development in phonological awareness (Anthony & Francis, 2005)--namely, isolation, blending, segmentation, deletion, and substitution.

There is wide consensus that universal screeners should include tasks that require students to isolate, blend, and segment phonemes in the early grades (e.g., International Dyslexia Association, 2022). As students continue to develop phonological awareness skills, it becomes useful to assess more advanced phonological awareness skills that require the manipulation of phonemes, such as deletion and substitution (Kilpatrick, 2015). Research has shown that these advanced phoneme manipulation tasks are some of the strongest predictors of decoding proficiency (e.g., Catts et. al., 2001). Currently, many commercially available screeners do not assess these more advanced phonological awareness skills, but the Literably Screener includes these advanced skills and can provide more sensitive data on students' levels of phonemic proficiency.

Phonological awareness is an oral and auditory skill, so the Literably Phonological Awareness subtest is an oral response task. Students are given the prompts one at a time and asked to provide answers verbally. In a segmentation task, for example, the student might be asked to "say moon," and then to "say the sounds you hear in moon." Students have 10 seconds to provide a response. The assessment is computer-administered, and student responses are recorded by Literably in the form of audio files and scored.

Phonics

The Literably Screener Phonics subtest, administered in grades K-3, assesses letter-name and letter-sound knowledge, along with the decoding of nonwords and real words. Research strongly supports screening for these alphabetic and decoding skills in the early grades (Fuchs et al., 2004).

Knowledge of letter names in the early grades strongly predicts later reading performance (Adams, 1990; Badian, 1995; Walsh et al., 1988), and knowledge of letter sounds is essential to the alphabetic principle and the development of decoding skills. There is wide consensus among researchers that universal screeners should measure letter name and letter sound knowledge in the early grades (e.g., National Center on Response to Intervention, 2013). The Phonics subtest includes letter name and letter sound items at Grades K-1 to assess students' acquisition of these prerequisite phonics skills.

Instead of requiring students to complete an inventory of all 26 letter names and sounds, the Literably Screener Phonics subtest requires students to provide 4 letter names and 4 letter

sounds (8 total) at the kindergarten level and 2 letter names and 2 letter sounds (4 total) in grade 1. The letters were selected based on their order of introduction in common phonics curricula. This approach saves screening time while providing data on students' levels of letter name and sound mastery. The letter sound items use only lowercase letters to clearly distinguish them from the letter name items, which use only uppercase letters.

Word reading is an essential component of screening batteries, as the skill is directly related to the ability to read more complex text. The Phonics subtest's real-word items measure students' ability to apply their knowledge of letter-sound relationships to the decoding and recognition of increasingly complex words. Because word reading has high utility for identifying children in need of intensive instruction (Fuchs, Fuchs, & Compton, 2004), this task comprises the majority of the Phonics subtest items.

The Phonics subtest also includes nonsense words or "pseudowords" in kindergarten and first grade in order to assess proficiency and automaticity with phonics rules. These make-believe words follow grade-appropriate phonics patterns and assess students' ability to apply their knowledge of grapheme-phoneme correspondences to inherently unfamiliar words. Nonword reading proficiency is a strong indicator of overall reading performance (Rathvon, 2004), including for English Language Learners (Vanderwood, Mike L., et al., 2008). Nonwords only use spelling patterns that could naturally occur in real words, and nonwords that sound like real words were excluded.

For the Phonics subtest, Literably content experts reviewed developmentally appropriate phonics skill progressions (e.g., CCSS, Blevins, etc.) along with the scopes and sequences from common phonics curricula used in grade schools in the U.S. (e.g., CKLA, Fundations, etc.). Based on an analysis of these skill progressions, content experts determined the appropriate skill coverage at each grade that would most match the phonics skills typically taught in grades K-5, and developed an item pool of nonwords and high-frequency decodable real words. Real words were selected to be grade-level appropriate based on published data on their age of acquisition, and forms were revised based on item statistics in order to maximize predictive value.

Table 5 below shows the skills covered on the Phonics subtest at each grade. The content blueprint within each grade is the same across seasons.

Table 5. *Phonics Test Blueprint*

Grade	**Item Type (Number of Items)**
Kindergarten (20 total items)	Letter sounds (4) Letter names (4) *Real Words:* ● CVC (2) ● VCe (1) ● Initial Blends (1) ● Initial Digraphs (1) ● Vowel Digraphs (1) ● Diphthongs (1) ● R-Controlled (1) *Nonwords:* ● Blends (1)

	● Consonant Digraphs (1) ● Vowel Digraphs (1) ● R-Controlled (1)
First Grade (25 total items)	Letter sounds (2) Letter names (2) *Real Words:* ● CVC (1) ● VCe (1) ● Initial Blends (1) ● Initial Digraphs (1) ● Vowel Digraphs (2) ● Diphthongs (2) ● R-Controlled (2) ● Inflectional Endings (1) ● Affixes (1) ● 2 Syllable with c-le (1) ● 2 Syllable with closed (1) ● 2 Syllable with open (1) ● 2 Syllable with VCe (1) ● 2 Syllable with vowel team (1) *Nonwords:* ● CVC (1) ● Consonant Digraphs (1) ● VCe (1) ● Vowel Digraph (1)
Second Grade (20 total items)	*Real Words:* ● CVC (1) ● VCe (1) ● Initial Blends (1) ● Initial Digraphs (1) ● Vowel Digraphs (3) ● Diphthongs (3) ● R-Controlled (2) ● Inflectional Endings (2) ● Affixes (1) ● 2 Syllable with c-le (1) ● 2 Syllable with closed (1) ● 2 Syllable with open (1) ● 2 Syllable with VCe (1) ● 2 Syllable with vowel team (1)
Third Grade (20 total items)	*Real Words:* ● CVC (1) ● VCe (1) ● Initial Blends (1) ● Initial Digraphs (1) ● Vowel Digraphs (1) ● Diphthongs (1)

	• Inflectional Endings (2) • Affixes (1) • 2 Syllable with c-le (2) • 2 Syllable with closed (2) • 2 Syllable with open (2) • 2 Syllable with VCe (1) • 2 Syllable with vowel team (1) • 3 Syllable with c-le (1) • 3 Syllable with open (1) • 3 Syllable with r-controlled (1)

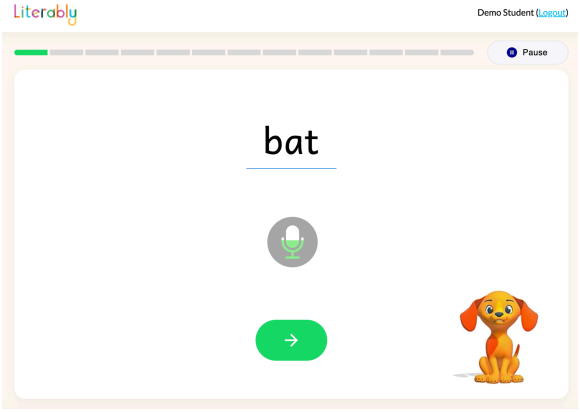
The Phonics subtest is a computer-administered task. In kindergarten and first grade, the assessment asks students to provide the names and sounds of letters and read a series of real and nonwords on the screen. In grades 2-3, students are only asked to read real words. Words and letters are presented individually, and students have 10 seconds to provide an oral response. Student responses are recorded by Literably in the form of audio files and scored.

Students hear the following prompts, based on the task:

- Letter Names: What is the name of this letter?
- Letter Sounds: What sound does this letter make?
- Real Word Reading: Read the word on the screen out loud.
- Non-Word Reading: Do your best to read this make-believe word.

Figure 1 shows a sample phonics item.

Figure 1. *Phonics - Sample Item*



Spelling

The Literably Screener Spelling subtest provides an indicator of students' ability to apply their knowledge of grapheme-phoneme correspondences to the production of written words. It is well-documented in the research literature that reading disorders such as dyslexia can be accompanied by deficits in encoding, and that the development of spelling skills contributes to reading ability (Moats, 2005). Katzir et al. (2006) found that spelling measures helped predict eventual reading proficiency, even after controlling for word reading proficiency, and research has demonstrated the appropriateness of spelling as a general outcome measure (Hosp & Hosp, 2003). A spelling inventory can reveal students' level of mastery of orthographic representations and help identify students who are at risk of reading difficulties. The Literably

Screener excludes the Spelling subtest at Kindergarten when many students are still gaining experience with encoding as well as the typing skills necessary for a digital spelling assessment.

Because encoding relies on the same underlying grapheme-phoneme correspondences students need to access in order to decode words for fluent reading (Moats, 2005), the Spelling subtest was designed to measure students' spelling ability along the same skill progression as the Phonics subtest. To this end, the sources consulted to produce the Spelling skill sequence were similar to those used to determine the Phonics skill progression.

Only regularly spelled words were included. Words were selected to be grade-level appropriate based on published data on their age of acquisition, and forms were revised based on item statistics in order to maximize predictive value. Words that can be interpreted as homophones were not included, and no words that appear on the Phonics subtests appear on the Spelling subtest, to avoid prior recent exposure to the word.

Table 6 below shows the skills covered on the Spelling subtest at each grade. The content blueprint within each grade is the same across seasons.

Table 6. *Spelling Test Blueprint*

Grade	**Item Type (Number of Items)**
First Grade (20 total items)	CVC (3) VCe (3) Initial Blends (1) Final Blends (1) Blends - CCVCC (1) Initial Digraphs (1) Final Digraphs (1) Digraphs - CCVCC (1) Vowel Digraphs (1) Diphthongs (1) R-Controlled (1) Inflectional Endings (1) 2 Syllable with c-le (1) 2 Syllable with closed (1) 2 Syllable with open (1) 2 Syllable with VCe (1)
Second Grade (20 total items)	CVC (1) VCe (2) Initial Blends (1) Final Blends (1) Initial Digraphs (1) Final Digraphs (1) Vowel Digraphs (1) Diphthongs (1) R-Controlled (2) Inflectional Endings (3) Affixes (1)

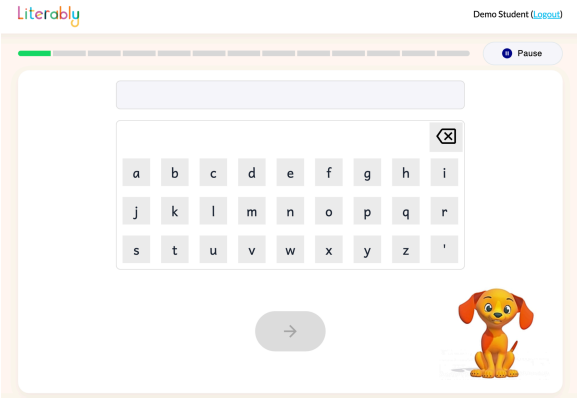
	2 Syllable with c-le (1) 2 Syllable with closed (1) 2 Syllable with open (1) 2 Syllable with VCe (1) 2 Syllable with vowel team (1)
Third Grade (20 total items)	VCe (1) Initial Blends (1) Initial Digraphs (1) Diphthongs (1) R-Controlled (1) Inflectional Endings (2) Affixes (1) 2 Syllable with c-le (2) 2 Syllable with closed (1) 2 Syllable with open (2) 2 Syllable with VCe (1) 2 Syllable with vowel team (2) 2 Syllable with r-controlled (1) 3 Syllable with c-le (1) 3 Syllable with open (1) 3 Syllable with VCe (1)
Fourth Grade (20 total items)	VCe (1) Initial Blends (1) Initial Digraphs (1) Inflectional Endings (3) Affixes (3) 2 Syllable with c-le (1) 2 Syllable with closed (1) 2 Syllable with open (1) 2 Syllable with VCe (1) 2 Syllable with vowel team (2) 2 Syllable with r-controlled (1) 3 Syllable with c-le (1) 3 Syllable with open (1) 3 Syllable with closed (1) 3 Syllable with VCe (1)
Fifth Grade (20 total items)	VCe (1) R-controlled (1) Inflectional Endings (3) Affixes (3) 2 Syllable with c-le (1) 2 Syllable with closed (1) 2 Syllable with open (1) 2 Syllable with VCe (1) 2 Syllable with vowel team (1) 3 Syllable with c-le (1) 3 Syllable with open (1)

	3 Syllable with closed (1) 3 Syllable with VCe (1) 4 Syllables with c-le (1) 4 Syllables with open (1) 4 Syllables with vowel team (1)

The Spelling subtest is computer-administered and can be given in a group setting. Each spelling word is dictated, spoken in the context of a short sentence, and repeated. For example, students might hear: “Porch. There is a swing on our front porch. Spell ‘porch.’” This format mirrors common paper-pencil spelling inventories, with the advantage that students’ responses are scored automatically by Literably. Students have the option of spelling the word using the on-screen keyboard or the keyboard on their device (e.g., iPad or computer). The onscreen keyboard supports younger students who may not have proficient typing skills. The Spelling assessment is not timed.

Figure 2 shows a sample Spelling item.

Figure 2. *Spelling - Sample Item*



Oral Reading Fluency

Ever since Deno et. al (1980) demonstrated the remarkable correlation between reading rate and achievement on standardized reading tests, a large body of research has shown oral reading fluency to be a particularly effective tool for identifying struggling readers and measuring their progress. This effectiveness is confirmed in the technical documentation accompanying widely-used oral reading assessments, such as the Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminski 2002) and AIMSweb (Howe & Shinn 2002). Like these well-known instruments, the Literably Screener Oral Reading Fluency (ORF) subtest measures the number of words a student reads out loud correctly in one minute. This words correct per minute (WCPM) score has demonstrated validity for the purpose of screening elementary school students for reading difficulties (Kilgus et al., 2014). In line with research and guidance on appropriate constructs to include in screening by grade (e.g., National Center on Response to Intervention, 2013), Literably begins screening for ORF in Grade 1, when students are expected to read grade-appropriate connected text with sufficient fluency and accuracy to support comprehension.

The Oral Reading Fluency (ORF) subtest, administered in grades 1-5, measures students’ ability to accurately read connected text for meaning. The task requires students to read one

short grade-level passage out loud and answer five multiple choice comprehension questions. ORF is a computer-administered task: the reading passage is presented to the student on screen, and the student's oral reading is recorded in the form of an audio file and scored for accuracy in order to produce a Words Correct Per Minute (WCPM) measure. Errors such as substitutions and omissions are marked on the graded assessment. A comprehension score provides instructionally valuable diagnostic information, but is not used to determine overall reading risk for the purposes of the screener.

To promote content validity, Literably's oral reading passages were selected to resemble the materials that students read in the classroom. Thus, Literably draws its reading passages, with permission, from children's trade books. Pictures are available for 1-3 passages. A team of elementary educators selected the excerpts from publishers' catalogs. Every team member had experience either as a teacher of relevant grades or as a reading specialist. Literably researchers then reviewed each proposed passage along the dimensions of age-appropriateness, bias, background knowledge required, and syntactic and phonic difficulty. Several school and district administrators experienced with assessment selection assisted with the process. When necessary, Literably content experts made small adjustments to texts to improve their conformity along these dimensions. Texts were also analyzed for passage statistics to maximize predictive validity.

Oral reading fluency is a well-established indicator of reading competency, as students who read connected texts fluently are more equipped to comprehend what they are reading. The method of measuring oral reading fluency whereby students read aloud, and a WCPM score is calculated based on a sample of the student's reading, is a widely implemented tool for identifying students in need of further reading intervention. Literably's ORF subtest replicates this gold standard of assessing ORF, with the added benefit of using technology to save significant time for teachers.

Vocabulary

Vocabulary knowledge is a key strand of Scarborough's Reading Rope and supports reading comprehension (Stahl & Fairbanks, 1986). In addition to being able to decode words, students need to understand their meanings in order to comprehend texts. Students with limited vocabulary knowledge may struggle with understanding instructions or performing reading-related tasks. The inclusion of the Vocabulary measure follows research indicating that the assessment of vocabulary is predictive of later reading outcomes (Sénéchal et al., 2006; Scarborough, 1998; Cunningham & Stanovich, 1997).

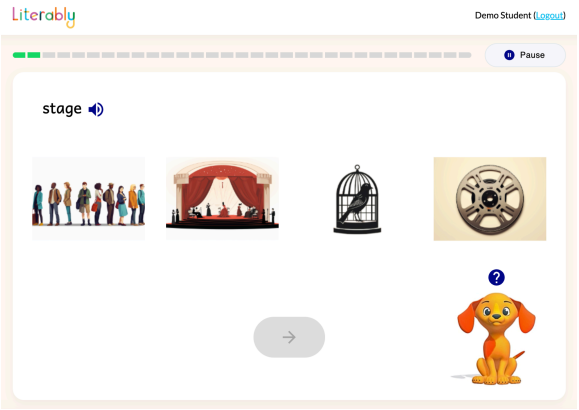
The Literably Screener Vocabulary subtest is a measure of students' receptive vocabulary knowledge. The Vocabulary assessment requires students to complete different types of developmentally appropriate tasks to gather data on different dimensions of vocabulary knowledge. The three task types are picture matching, synonym, and context. Picture matching and synonym tasks are used in Kindergarten, all three tasks are used in grades 1 and 2, and synonym and context questions are used in grades 3 and above. There are 20 questions on each form.

Words in the Vocabulary assessment include high-frequency, high-utility words that students should be expected to know in order to comprehend texts in academic settings. Starting in grade 1, the assessment also includes content-specific words across Language Arts, Math, Science, and Social Studies domains. Words were selected from well-established vocabulary lists compiled to guide reading instruction in the early grades (e.g., Biemiller, 2009; Marzano &

Simms, 2013), using age of acquisition to align words to appropriate grade levels. Whenever possible, words were selected to represent diversity in part of speech, topic, and academic domain.

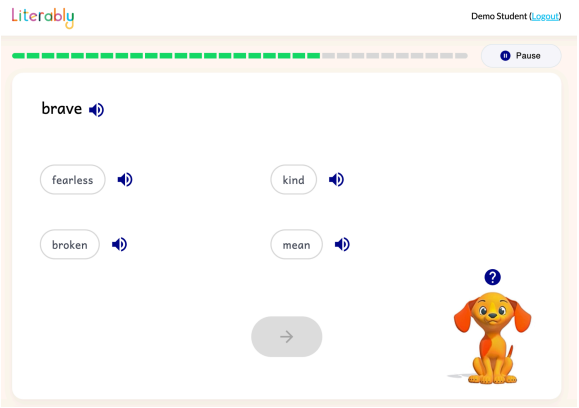
The picture matching task asks students to identify the picture that best matches the target word from four pictures. Students can read the word independently or have it read aloud. This picture-vocabulary format is used in well-researched assessments of receptive vocabulary, such as the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007). Target words for these tasks were selected to be concrete and visualizable.

Figure 3. *Vocabulary - Sample Picture-Matching Item*



The synonym task asks students to select the word that is most similar in meaning to the target word, using the prompt “Select the word that has the same or similar meaning.” The task requires students to demonstrate understanding of word meaning in the absence of contextual clues. Students may read the words or have them read aloud. Distractor answer choices may include words that are related semantically but not synonyms, as well as words that resemble the target word in sound or spelling.

Figure 4. *Vocabulary - Sample Synonym Item*



The context, or fill-in-the-blank task, introduced in grade 1, asks students to demonstrate understanding of words in the context of a sentence. Knowing word meaning in context most closely approximates skills students need in order to achieve passage comprehension. Students select among the four answer choices after hearing the prompt “Select the word that best

completes the sentence." To ensure age appropriateness, sentences at Grades K-1 are capped at 10 words, sentences at Grades 2-3 are capped at 12 words, and sentences at Grades 4-5 are capped at 14 words. Whenever possible, the answer blanks appear near the end of the sentence to minimize the need to reread the sentence.

Students have the option of reading the sentences and answer choices themselves, or having them read aloud by the device. Distractor answer choices may include words that resemble the word in meaning but do not fit in the context of the sentence, or words that resemble the target word in sound or spelling. Incorrect answer choices were also selected to be grade-appropriate and have either the same or lower age of acquisition as the target word.

Figure 5. *Vocabulary - Sample Context Item*

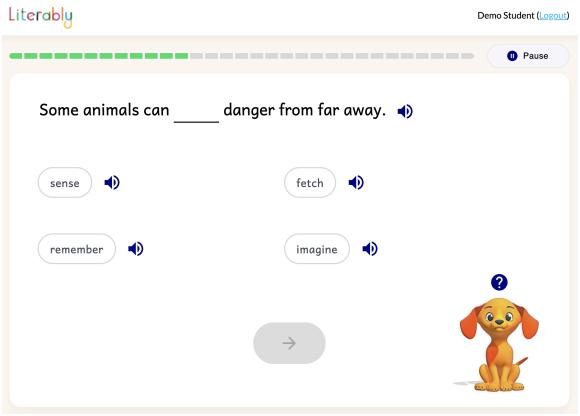


Table 7 below shows the distribution of question types on the Vocabulary subtest at each grade. The content blueprint within each grade is the same across seasons.

Table 7. *Vocabulary Test Blueprint*

Grade	**Item Type (Number of Items)**
Kindergarten (20 total items)	Picture Vocabulary (5) Synonyms (15)
First Grade (20 total items)	Picture Vocabulary (7) Synonyms (6) Fill-in-the-blank (7)
Second Grade (20 total items)	Picture Vocabulary (9) Synonyms (6) Fill-in-the-blank (5)
Third Grade (20 total items)	Synonyms (5) Fill-in-the-blank (15)
Fourth Grade (20 total items)	Synonyms (5) Fill-in-the-blank (15)
Fifth Grade	Synonyms (5)

(20 total items)	Fill-in-the-blank (15)

Reading Comprehension

In the Simple View of Reading model (Gough & Tunmer, 1986), reading comprehension is the product of decoding ability and language comprehension. Research supports the screening of passage reading comprehension starting at grade 2, as students apply their decoding and language comprehension skills to reading and understanding connected text (Torgesen, 2002; Jenkins et al., 2007).

The Reading Comprehension subtest, administered in grades 2-5, measures a student's ability to read grade-appropriate literary and informational texts and demonstrate understanding of the text. Along with the Oral Reading Fluency subtest, Reading Comprehension can provide crucial data regarding students' ability to read and comprehend connected text at their grade level.

The Reading Comprehension assessment is computer-administered and can be given whole-class. The format of the Reading Comprehension subtest mirrors authentic passage-and-question tasks that students often complete in the classroom or on standardized assessments. Each Reading Comprehension assessment contains two passages that students read silently: one informational (non-fiction) and one literary (fiction). Texts are not accompanied by pictures. Each text is followed by six multiple-choice comprehension questions, presented one at a time. Students are able to refer back to the text when answering the questions, and they can go back to previous questions within each text. However, they cannot navigate back to a previous text once they have moved to the next text. The assessment is untimed.

The Reading Comprehension passages were developed by extracting passages or entire texts from authentic works of children's literature and nonfiction works. Flesch-Kincaid readability estimates were used to determine grade-appropriateness of the passages.

Literably content experts worked with a team of experienced educators to write the comprehension questions and then revised them based on feedback from a broader pool of educators. To maximize alignment with the curricula of schools in the United States, the authors wrote every multiple-choice comprehension question to assess one or more of the Common Core State Standards (Common Core State Standards Initiative, 2010). The questions comprise a mix of literal and inferential questions. Questions were also written to be dependent on the text and only answerable using information from the text.

Figure 6. *Reading Comprehension - Sample Passage*

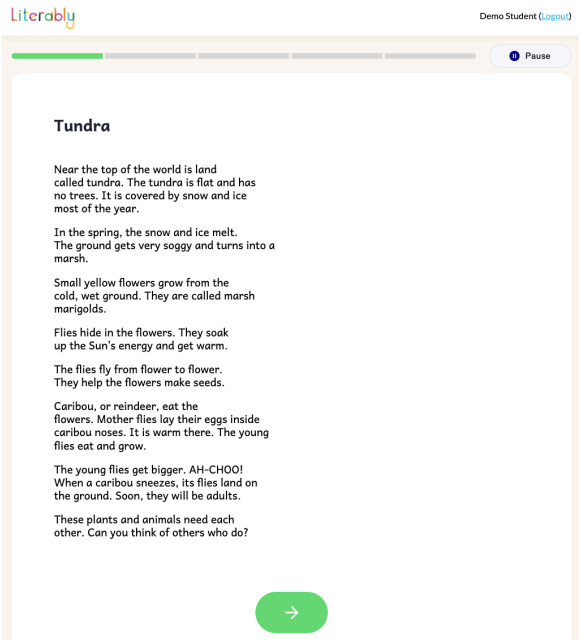
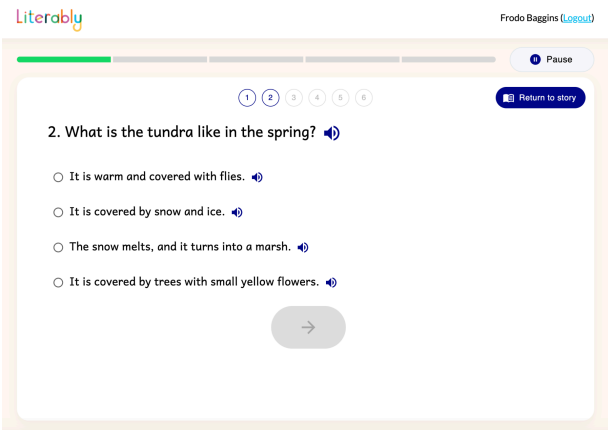


Figure 7. *Reading Comprehension - Sample Question*



Scoring Method

Scoring Method

The Literably Screener is scored by Literably, and test administrators are not responsible for scoring any portion of the assessment.

At each benchmark assessment period, the Literably Screener produces two types of scores:

1) A raw score for each subtest
2) A composite score

Table 8 shows how each Literably Screener subtest is scored and the outcomes reported.

Table 8. *Scoring method and reported outcomes by subtest*

Subtest	Response type	Scoring Method	Reported outcome(s)	Scoring Entity and Time
RAN	Oral response	Transcription	Average items correct per minute; duration in seconds	Scoring is done by trained Literably graders within 24 hours.
Phonological Awareness	Oral response	Transcription	Number correct out of number attempted	Scoring is done by trained Literably graders within 24 hours.
Phonics	Oral response	Transcription	Number correct out of number attempted	Scoring is done by trained Literably graders within 24 hours.
Spelling	Typed response	Dichotomously scored at the item level	Number correct out of number attempted	Scoring is done immediately and automatically by Literably.
Oral Reading Fluency	Oral response, timed	Transcription	Words correct per minute (WCPM); percentage accuracy	Scoring is done by a combination of trained Literably graders and ASR within 24 hours.
Vocabulary	Selected response	Dichotomously scored at the item level	Number correct out of number attempted	Scoring is done immediately and automatically by Literably.
Reading Comprehension	Selected response	Dichotomously scored at the item level	Number correct out of number attempted	Scoring is done immediately and automatically by Literably.

Transcription

Four of the Literably Screener subtests (Phonological Awareness, Phonics, RAN, and Oral Reading Fluency) measure foundational literacy skills that are best assessed orally. Historically, because computers cannot process speech with human accuracy, schools have faced a difficult choice when trying to assess these skills: either require teachers to manually administer one-on-one assessments, thereby consuming hours of valuable instructional time; or rely on computer-administered assessments that use selected-response items as indirect and imperfect measures of these skills, thereby reducing the face validity of the data. More recently, some computer-administered assessments have begun using automatic speech recognition (ASR) to measure these skills, but ASR still cannot match human accuracy, and it especially struggles with noisy classroom audio, as well as varied articulation and dialect.

To score the Phonics, Phonological Awareness, and RAN subtest, Literably uses normed human transcription. Literably's transcriptionists are selected using a competitive application process wherein successful applicants must transcribe a sample of student audio recordings with a high rate of concordance with the transcriptions of Literably staff members. Once hired, transcriptionists regularly face random test recordings to ensure their transcription accuracy remains high.

In order to ensure equitable scoring of oral responses for students with varied articulation and/or dialect, Literably's human transcribers are trained to accept diverse pronunciations. As further protection against inequitable scoring, Literably's scoring engine includes an automatic check that scans the final transcript for "errors" that reflect common alternative pronunciations based on articulation and/or dialect. If any of these "errors" are found, the Literably scoring engine will automatically override the error. This automatic check includes rules that apply in general to all Literably students, as well as rules that are specifically applied based on the student's home language. This supports the accurate scoring of oral responses for students with speech differences, English language learners, and very young students.

For Oral Reading Fluency, Literably uses a "human in the loop" approach that blends custom ASR and normed human transcription to deliver human-level accuracy at an acceptable cost. When Literably receives an audio recording, it slices the recording into segments, gauges the difficulty of each segment with ASR, and then flags the segment for ASR or human transcription. Once the segments have been transcribed, Literably rejoins them into one transcript and presents the complete transcript to the teacher. This approach allows Literably to directly and accurately assess oral reading fluency.

RAN Scoring

As students complete the Literably RAN subtest, their responses are recorded, and the audio file(s) are uploaded to Literably for scoring. Literably automatically computes and reports the duration of the recordings. Literably uses normed human transcription to transcribe the audio files, and the transcripts are matched to the stimulus arrays as annotations, so teachers can see where the students deviated from the stimulus items. Teachers receive results within 24 hours.

In addition to the duration, the Literably RAN subtest also generates an average items correct per minute score. This is an average of the items correct per minute scores on the two RAN tasks: numbers and letters. The average items correct per minute score is used to calculate the overall Literably Screener composite score.

Phonological Awareness and Phonics Scoring

As students complete the Phonological Awareness and Phonics subtests, their responses are recorded, and the audio files are uploaded to Literably for scoring. The audio files are transcribed using normed human transcription, and the transcripts are matched to the items as annotations, so teachers can see the students' exact responses. Each item is scored as correct or incorrect, and teachers receive results within 24 hours.

Each subtest reports a raw score indicating the number of items correct. The raw score is used to calculate the overall Literably Screener composite score.

Table 9 lists specific scoring rules that apply to these subtests.

***Table 9**. Phonological Awareness and Phonics Scoring Rules*

Scoring rule	Phonological Awareness	Phonics: Letter names	Phonics: Letter sounds	Word / Nonword Reading
Self-corrections are marked correct.	x	x	x	x
Incorrect responses are transcribed, and the item is marked with a slash.	x	x	x	x
Omissions are marked incorrect and grayed out.	x	x	x	x
Insertions before or after a correct response do not count as errors.	x	x	x	x
Schwa sounds added to consonants do not count as errors.	x		x	
Only words that are completely segmented are marked correct.	x			
Sound elongation is permitted.	x			
Words “sounded out” are incorrect, unless also blended.				x

Spelling Scoring

As students submit their typed responses on the Spelling subtest, their responses are uploaded to Literably for scoring. Literably automatically compares the student’s response with the correct response, and the items are dichotomously scored as correct or incorrect. Teachers receive results immediately after the assessment is completed.

The Spelling subtest reports a raw score indicating the number of items correct. The raw score is used to calculate the overall Literably Screener composite score.

Oral Reading Fluency Scoring

As students read the text aloud, the audio file is streamed to Literably for scoring. Literably has the student audio transcribed using a blend of automatic speech recognition (ASR) and normed human transcription, and the transcripts are matched to the text as annotations, so teachers can see exactly how students deviated from the text. Each word is scored as correct or incorrect.

Literably’s scoring engine analyzes every oral response using automatic speech recognition (ASR) and allocates audio segments to ASR or human transcription in order to maximize transcription accuracy while controlling transcription cost. All scores are reported within 24 hours of the student’s submission.

The Oral Reading Fluency subtest generates 2 scores: words correct per minute (WCPM) and percentage accuracy. Words correct per minute is the number of words read correctly per minute of oral reading. Percentage accuracy is calculated as the total words read correctly

divided by the total words attempted, multiplied by 100. Both scores are automatically calculated by Literably. The WCPM score is used to calculate the Literably Screener composite score.

Literably oral reading adheres to the following scoring rules:

- Omissions and substitutions are counted as errors.
- Insertions, repetitions, and self-corrections are not counted as errors.

The Oral Reading Fluency subtest also generates a comprehension score; however, this score is not used for calculating the composite score for the screener.

Vocabulary and Reading Comprehension Scoring

As students complete the Vocabulary and Reading Comprehension subtests, their responses are uploaded to Literably for scoring. All items are multiple choice, and the items are dichotomously scored as correct or incorrect.

Each subtest reports a raw score indicating the number of items correct. The raw score is used to generate the overall Literably Screener composite score.

Literably Screener Composite Score

For the purpose of interpreting results from the Literably Screener, teachers should attend primarily to the composite score, which is derived from the results on the subtests completed during each testing window. The composite score is the best overall predictor of later reading outcomes and takes into account all domains that are crucial to reading success at each grade level.

The Literably Screener composite score enables cut score analysis and performance classifications at the whole test level. The composite score alone is not inherently meaningful to educators. However, the performance classifications that are enabled by the composite score have clear instructional relevance for educators. Literably Screener cut scores were established to optimize classification accuracy relative to an external end-of-year measure of reading proficiency and achievement.

For easy interpretation, Literably Screener composite scores are translated into performance classifications and color-coded based on the Literably cut scores. Later sections of this technical manual describe the methodology for determining cut scores and the interpretation of screener performance classifications.

Other Literably Screener Data

Beyond the scores described in the sections above, Literably also provides data related to how students are performing on the discrete skills covered by the subtests on the Literably Screener. While these skill data do not factor into the composite score and evaluation of risk, they may be useful for gaining a closer understanding of students' strengths and difficulties in various domains, for the purposes of informing future interventions and supports.

Special Scoring Considerations

The Literably Screener is appropriate for the large majority of K-5 students learning to read in English, for the primary purpose of identifying students who may be at risk of developing

reading difficulties, including students with disabilities, English learners, and gifted students. Some additional considerations are described below, and the Bias Analyses section of this technical manual includes psychometric data related to the validity of the instrument for various subgroups.

English Language Learners

The Literably Screener is appropriate for assessing English Learners (ELs). Research suggests that ELs acquire English literacy skills in a similar manner to native English-speaking children, and that early identification for reading risk is beneficial for ELs as well as native speakers (Chiappe, Siegel, & Wade-Woolley, 2002, Lesaux & Siegel, 2003).

The oral reading of ELs often reveals interference from a student's native language (Price et al., 2009). Students with limited knowledge of the English language will frequently make use of phonological, orthographic, or morphological patterns from their native languages in their English reading. For this reason, most teacher-administered oral reading assessments include scoring rules instructing teachers to forgive these errors. For example, the Administration and Scoring Guide of the Dynamic Indicators of Basic Literacy Skills (DIBELS) states: "The student is not penalized for imperfect pronunciation due to dialect, articulation, or second language interference" (Good & Kaminski, 2014, p. 19).

To mitigate this interference, Literably applies home-language-specific lists of acceptable phonemic variations that are more likely to be driven by a student's language status than her decoding abilities. In determining the contents of this list, our process considers: (1) analysis of recordings of oral reading by EL and non-EL students in Literably's dataset, (2) user feedback, and (3) well-known pronunciation errors for ELs as described by experts in bilingual education. For oral subtests (ORF, Phonics, RAN, and Phonological Awareness), Literably maintains and implements lists of acceptable variations for the following languages: Spanish, Arabic, Mandarin, Somali, Russian, Hmong, Vietnamese, Korean, Tagalog, Haitian Creole, French, Hebrew. After a student's response is transcribed, Literably's software automatically checks for and overrides any errors that are considered acceptable phonemic variations.

Very Young Students

The Literably Screener is designed to be used by students in kindergarten through second grade. The student interface is easy to navigate, and all subtests except for ORF and Reading Comprehension feature a friendly animated character (*Figure 8*) that is engaging to younger users. Students do not need any prior education or classroom experience to use Literably, and Literably offers embedded features that allow teachers to model for students how to navigate the student assessment platform.

Figure 8. *Literably Animated Avatar: Rudy*

On iPads and tablets, the Literably student interface allows students to use touchscreen technology to navigate through the assessment, without the need for typing on a keyboard or using a mouse.

The Literably Screener can be administered in a group setting or individually with students. For all students and especially the youngest students (e.g., students in early kindergarten or who are younger than typical kindergarten age), if a student needs additional assistance navigating the platform, teachers have the option to assess the student one-on-one and provide navigation assistance.

Disability Status

The Literably Screener is appropriate for use with students receiving special education services. The Test Administration section of this manual describes accommodations that can be implemented with Literably to support students with IEPs or 504 plans.

There are some student populations for whom the Literably Screener may not be appropriate:

1. For students with significant hearing impairments, the Phonological Awareness subtests may not be an appropriate measure, since the assessment is an auditory task that requires the isolation and manipulation of sounds.
2. For students with significant speech impairments that affect articulation, the Phonics, Phonological Awareness, and Oral Reading Fluency assessments may produce lower-than-expected scores. For students with speech impairments that affect fluency, the Oral Reading Fluency subtests may produce lower-than-expected scores. Literably recommends the use of professional judgment or consultation with a speech therapist when deciding whether to administer the Literably Screener to students with speech disabilities.
3. Finally, the Literably Screener may not be appropriate for students with very severe disabilities, such as nonverbal students, for whom reading connected text is not an IEP goal. For these students, alternative assessments may provide more meaningful information on progress towards individualized goals and objectives.

It is recommended that district personnel review the above considerations along with the available accessibility features described in the Administration Manual in order to provide local guidance related to the administration of the Literably Screener for these populations.

Reliability

Subtest Reliabilities

The reliabilities of the Literably Screener subtests were estimated using test-retest and coefficient alpha.

For the non-item-based measures - rapid automatized naming (RAN) and oral reading fluency (ORF) - Literably and its research partners evaluated test-retest reliability. Participating students completed the same RAN and ORF forms twice in the same assessment window, and researchers correlated the scores from both test administrations. A correlation coefficient above 0.70 represents acceptable reliability, above 0.80 represents good reliability, and above 0.90 represents excellent reliability. *Table 10* presents the reliability estimates for RAN and ORF. The RAN estimates reflect the reliability of the Literably Screener RAN subtest, which includes Letters and Numbers. RAN and ORF demonstrate good to excellent reliability, with all coefficients exceeding 0.80.

Table 10. *Test-Retest Reliability for Literably RAN and ORF*

Grade	RAN		ORF	
	N	r	N	r
K	124	0.84	-	-
1	230	0.80	234	0.89
2	-	-	195	0.92
3	-	-	396	0.83
4	-	-	279	0.88
5	-	-	284	0.91

For the remaining measures, including Phonological Awareness (PA), Phonics, Spelling, Vocabulary and Comprehension, Literably and its research partners calculated coefficient alpha. Coefficient alpha (Cronbach, 1951) is a measure of internal consistency–the extent to which the items on an instrument are related. Alpha values above 0.60 are generally considered acceptable, and values above 0.80 are considered very good.

Table 11 presents internal consistency reliability estimates for PA, Phonics, Spelling, Vocabulary and Comprehension. Overall, these subtests show acceptable to excellent reliability–with all but 1 estimate exceeding 0.6, and the large majority exceeding 0.8. PA, Phonics, Spelling and Vocabulary show very strong reliability, with all but 2 estimates exceeding 0.8 and those 2 falling at 0.79. Comprehension's reliability is merely acceptable. Lower internal consistency reliability for the Comprehension subtest is expected in that it includes two sets of passage-dependent questions addressing two different texts.

Table 11. *Internal Consistency Reliability for PA, Phonics, Spelling, Vocab, Comprehension*

Subtest	Form	K		1		2		3		4		5	
		N	α	N	α	N	α	N	α	N	α	N	α
PA	A	129	0.86	213	0.84	-	-	-	-	-	-	-	-
	B	140	0.84	207	0.84	-	-	-	-	-	-	-	-
	C	145	0.83	205	0.83	-	-	-	-	-	-	-	-
Phonics	A	144	0.91	215	0.91	184	0.92	239	0.92	-	-	-	-
	B	148	0.89	215	0.92	179	0.93	236	0.94	-	-	-	-
	C	146	0.85	214	0.91	171	0.93	238	0.93	-	-	-	-
Spelling	A	-	-	217	0.91	182	0.91	251	0.91	295	0.90	294	0.92
	B	-	-	227	0.91	184	0.92	250	0.91	298	0.89	296	0.92
	C	-	-	216	0.91	179	0.92	257	0.93	301	0.92	294	0.90
Vocab	A	159	0.85	224	0.82	180	0.87	246	0.86	292	0.86	294	0.80
	B	162	0.89	227	0.80	184	0.85	237	0.89	293	0.84	292	0.81
	C	167	0.86	221	0.79	185	0.80	255	0.84	295	0.80	289	0.79
Comprehension	A	-	-	-	-	183	0.74	241	0.79	280	0.76	281	0.75
	B	-	-	-	-	180	0.79	240	0.82	274	0.75	283	0.79
	C	-	-	-	-	173	0.75	242	0.80	277	0.72	281	0.78
	Fall	-	-	-	-	190	0.67	246	0.76	321	0.72	327	0.73
	Winter	-	-	-	-	170	0.72	243	0.76	252	0.70	259	0.54
	Spring	-	-	-	-	175	0.72	230	0.62	256	0.70	257	0.74

Composite Score Reliability

In addition to the subtest reliabilities, Literably researchers estimated the reliability of the Literably Screener Composite Score at each grade and season using Feldt & Brennan's (1989) stratified alpha method. Stratified alpha uses the reliabilities and variances for each subtest to estimate the reliability of the composite score using the equation below, where i is a subtest, σ^2 is variance, α is reliability, and x is the composite.

$$Stratified\,\alpha = 1 - \frac{\Sigma\sigma^2_i\,(1-\alpha_i)}{\sigma^2_x}$$

Table 12 reports the reliability of the Literably Screener Composite Score for each grade and season. Sample sizes are not reported because sample sizes vary by subtest. All reliability estimates exceed 0.8 and all but two exceed 0.9, providing strong evidence of the Literably Screener Composite Score's reliability.

Table 12. *Composite Score Reliability (Stratified Alpha) by grade and season*

Grade	Fall	Winter	Spring
K	0.92	0.90	0.89
1	0.94	0.93	0.93
2	0.95	0.95	0.95
3	0.90	0.90	0.89
4	0.92	0.91	0.91
5	0.94	0.94	0.93

Delayed Alternate-Form Reliability

The Literably Screener is a fixed form assessment - with forms defined for each grade and season. However, to examine form equivalence, Literably researchers analyzed delayed alternate-form reliability of the Literably Screener Composite Score fall to winter and winter to spring.

Table 13 shows the results of the delayed alternate-form reliability analyses. Delayed alternate-form reliability analyses are generally expected to yield more moderate coefficients due to the time elapsed between administrations. Nevertheless, the Literably Screener Composite Score demonstrated strong reliability. Of the 12 coefficients, 11 were good to excellent (above 0.8), and 1 fell at the very high end of acceptable (0.79).

Table 13. *Delayed Alternate-Form Reliability of the Literably Screener Composite Score*

Grade	Fall-Winter		Winter-Spring	
	N	r	N	r
K	43	0.87	37	0.89
1	74	0.86	94	0.82
2	107	0.90	118	0.85
3	148	0.84	153	0.88
4	178	0.83	170	0.79
5	198	0.84	192	0.85

Standard Error of Measurement

Literably researchers estimated the standard error of measurement (SEM) of the Literably Screener Composite Score for each grade and season. To calculate the SEM, Literably researchers multiplied the standard deviation by the square root of one minus the reliability. *Table 14* reports the Literably Screener Composite Score SEMs for Grades K-5.

Table 14. *Literably Screener Composite Score SEMs by Grade and Season*

Grade	Fall	Winter	Spring
K	7	12	11
1	12	18	19
2	9	12	11
3	14	15	16
4	12	12	12
5	11	11	12

Validity

Content Validity

Overview

The Literably Screener tasks authentically and directly measure the key constructs that contribute to reading success. *Table 15* shows the alignment between Literably Screener subtests and key constructs.

Table 15. *Literably Screener Subtests and Key Constructs*

Subtest	**Construct(s)**
Rapid Automatized Naming (RAN)	Rapid automatized naming
Phonological Awareness (PA)	Phonological and Phonemic Awareness
Phonics	Letter Naming Letter-Sound Correspondence Real Word Reading Nonword Reading
Spelling	Encoding
Oral Reading Fluency (ORF)	Oral Text Reading Accuracy and Rate
Vocabulary	Vocabulary (receptive)

Comprehension	Reading Comprehension

The sections below provide further detail on the content validity for each Literably Screener subtest.

Rapid Automatized Naming (RAN)
Grades Administered: K-1
Construct(s) Measured: Rapid automatized naming

The Literably RAN subtest measures students' ability to quickly name familiar items (letters and numbers) aloud, which is a strong predictor of reading growth (Lervåg & Hulme, 2009; Al Otaiba & Fuchs, 2002; Nelson et al., 2003). Literably RAN is computer-administered with standardized instructions. Students respond aloud, and their verbal responses are recorded as audio files and scored for accuracy and speed.

Literably developed Literably RAN based on an extensive review of the RAN literature, and Literably RAN closely mirrors best practices in RAN assessment. Students are presented with arrays composed of repeating, randomly-ordered groups of five familiar items (Georgiou et al., 2013; McWeeny et al, 2022). Each array contains five rows of ten items, and students name the items left to right and top to bottom. Literably test developers consulted with educators to select numbers (4, 1, 7, 8, 2) and letters (T, X, A, J, U) that students would be unlikely to confuse with each other. The stimulus items and example arrays for each RAN task are displayed in *Table 16.*

Table 16. *RAN Tasks, Stimulus Items and Sample Arrays*

Tasks	**Stimulus Items**	**Sample Arrays**
Letters	A, T, U, X, J	T X J U A T U A J X T X J A U T J X A U X A J U T A X T U J X T J A U T J X U A J T U A X J A U X T
Numbers	7, 1, 8, 2, 4	7 1 8 2 4 8 7 4 2 1 8 1 2 7 4 1 7 4 2 8 4 8 7 1 2 7 8 1 2 4 2 4 1 7 8 2 7 8 4 1 2 8 7 4 1 8 1 4 2 7

Literably developed, field-tested, and validated forms for all four of the typical RAN tasks: Objects, Colors, Letters and Numbers at Grades K-5. However, Literably and its research partners found that the alphanumeric subtests (Letters and Numbers) outperformed the non-alphanumeric subtests (Objects and Colors) in predictive validity, which replicated the finding of a recent meta-analysis (Araujo, Reis, Magnus Petersson, & Faisca, 2015). Therefore, to reduce testing time, the Literably Screener Composite draws from Letters and Numbers, and students are not required to complete Objects and Colors. Literably researchers also found that RAN meaningfully contributed to the Literably Screener's predictive validity at Grades K-1, but it

no longer meaningfully contributed at Grades 2-5. Therefore, to reduce testing time, RAN is included in the Literably Screener Composite at Grades K-1, and it is excluded at Grades 2-5.

Phonological Awareness (PA)
Grades Administered: K-1
Construct(s) Measured: Phonological and Phonemic Awareness

The Literably Phonological Awareness (PA) subtest directly measures students' ability to identify and manipulate sounds in spoken words in grades K-1. Literably PA is an untimed computer-administered assessment with standardized instructions. Phonological awareness is oral and auditory, so Literably PA is an oral response assessment. Literably presents audio prompts to students one at a time, and students respond aloud. Student responses are recorded as audio files and scored for accuracy.

To ensure developmentally-appropriate items, Literably test developers completed a literature review–including relevant research papers, standards, curricula and assessments–and developed grade-specific test blueprints that reflect the overlap of current and best practices (i.e., the skills that are taught and assessed, and the skills that should be taught and assessed at each grade). The test blueprints for each grade remain the same across seasons. *Table 17* presents the skills covered for Grades K-1. It is worth noting that Literably's literature review found broad consensus that screening assessments should include isolating, blending and segmenting phonemes. Literably researchers also found evidence that phoneme manipulation tasks, including deletion and substitution–while less popular among existing screeners–are strongly predictive of later reading proficiency (Catts et al., 2001; Kilpatrick, 2015). In the interest of providing a more complete and predictive assessment of students' phonological awareness, Literably PA asks students to isolate, blend, segment, delete and substitute, as reflected in *Table 17*.

Table 17. *Phonological Awareness - Skills Covered by Grade*

Grade	**Blueprint**
Kindergarten 20 items	Isolating initial, medial, and final phonemes Blending 3 and 4 phonemes Segmenting 3 and 4 phonemes Deleting compound words and initial phonemes Substituting initial, medial, and final phonemes
First Grade 20 items	Isolating initial, medial, and final phonemes Blending 3 and 4 phonemes Segmenting 3, 4, and 5 phonemes Deleting initial phonemes and final phonemes of initial blends Substituting initial, medial, and final phonemes

Once the test blueprints were established, Literably test developers wrote items aligned to the target skills. To ensure grade-appropriate items, words were selected based on age of acquisition (Kuperman et al., 2012), and each item was reviewed by multiple test developers. Test developers also ensured no overlap in words between the Phonological Awareness, Phonics and Spelling subtests to avoid priming the student and/or over-sampling a single word.

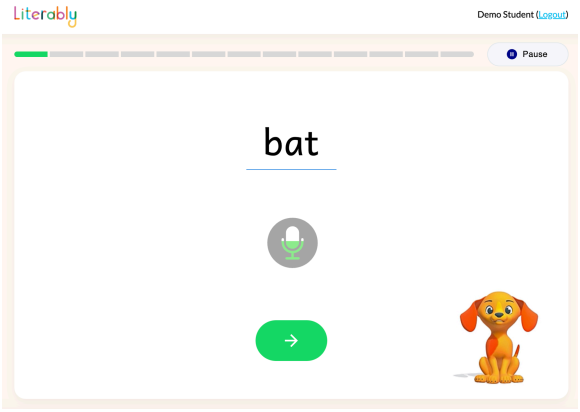
Finally, during the 2023-2024 psychometric study, Literably and its research partners analyzed Literably PA item quality using classical statistics (p-value, item-total correlation) and item response theory (IRT) statistics (e.g., difficulty parameters, DIF) to ensure items were appropriately challenging, unbiased and effective at distinguishing between levels of skill on the target construct. Sub-optimal items (i.e., p-values >0.95 or <0.05; item-total correlations < 0.2; IRT difficulty > 4 or < -4; absolute value of standardized p-DIF > 0.1; point-biserial values < 0.2; outfit or infit values > 1.5) were flagged for review by Literably test developers. All flagged items are either discarded or retained for further analysis.

Literably developed, field-tested and validated PA forms for Grades K-5. However, Literably and its research partners found that PA did not meaningfully contribute to the Literably Screener's predictive validity above Grade 1. Therefore, to reduce testing time, PA is included in the Literably Screener Composite at Grades K-1, and it is excluded at Grades 2-5.

Phonics
Grades Administered: K-3
Construct(s) Measured: Letter Naming, Letter-Sound Correspondence, Real Word Reading, Nonword Reading

The Literably Phonics subtest directly measures K–2 students' ability to identify letter names and letter sounds, as well as to read aloud phonically regular, grade-appropriate words and nonwords. Literably Phonics is an untimed computer-administered assessment with standardized instructions. Phonics is best assessed orally, so Literably is an oral response assessment. Literably presents an audio prompt to the student (e.g., What is the name of this letter?) followed by a visual stimulus (letter or word), and the student responds aloud. Student responses are recorded as audio files and scored for accuracy. *Figure 9* shows a sample phonics item.

Figure 9. *Phonics - Sample Item*



To ensure developmentally-appropriate items, Literably test developers completed a literature review–including relevant research papers, standards, curricula and assessments–and developed grade-specific test blueprints that reflect the overlap of current and best practices (i.e., the skills that are taught and assessed and the skills that should be taught and assessed at each grade). Literably's literature review found broad consensus that screening assessments should include word and pseudoword reading, as well as letter name and letter sound knowledge in the early grades (Fuchs, Fuchs, & Compton, 2004). The test blueprints for each grade remain the same across seasons. *Table 18* presents the skills covered for Grades K-3.

Table 18. *Phonics - Skills Covered by Grade*

Grade	Item Types
Kindergarten 20 items	Letter sounds Letter names *Real Words:* VC, CVC, VCe, blends, consonant digraphs, r-controlled, vowels, digraphs, diphthongs *Non-Words:* CVC, VCe, blends, consonant digraphs, r-controlled vowels, vowel digraphs, diphthongs
First Grade 25 items	Letter sounds Letter names *Real Words:* CVC, VCe, blends, consonant digraphs, r-controlled vowels, vowel digraphs, diphthongs, affixes, inflectional endings, 2 syllables words with closed, open, VCe, vowel team, c-le, and r-controlled syllables *Non-Words:* CVC, VCe, blends, consonant digraphs, r-controlled vowels, vowel digraphs, diphthongs
Second Grade 20 items	*Real Words:* CVC,VCe, blends, consonant digraphs, r-controlled vowels, vowel digraphs, diphthongs, affixes, inflectional endings, 2 and 3 syllable words with closed, open, VCe, vowel team, c-le, and r-controlled syllables
Third Grade 20 items	*Real Words:* CVC, VCe, blends, consonant digraphs, r-controlled vowels, vowel digraphs, diphthongs, affixes, inflectional endings, 2 and 3 syllable words with closed, open, VCe, vowel team, c-le, and r-controlled syllables

Once the test blueprints were set, Literably test developers wrote items aligned to the target skills. To ensure grade-appropriate items, words were selected based on age of acquisition (Kuperman et al., 2012), and each item was reviewed by multiple test developers. Nonwords were written to be consistent with English spelling, and nonwords that sound like real words were excluded. Test developers also ensured no overlap in words between the Phonological Awareness, Phonics and Spelling subtests to avoid priming the student and/or over-sampling a single word. For letter names and letter sounds, the Kindergarten blueprint includes 4 letter names and 4 letter sounds, and the Grade 1 blueprint includes 2 letter names and 2 letter sounds. The letters were selected based on their order of introduction in widely-used phonics curricula. The letter-sound items use only lowercase letters to clearly distinguish them from the letter name items, which use only uppercase letters.

Finally, during the 2023-2024 psychometric study, Literably and its research partners analyzed Literably Phonics item quality using classical statistics (p-value, item-total correlation) and item response theory (IRT) statistics (e.g. difficulty parameters, DIF) to ensure items were appropriately challenging, unbiased and effective at distinguishing between levels of skill on the target construct. Sub-optimal items (i.e. p-values >0.95 or <0.05; item-total correlations < 0.2; IRT difficulty > 4 or < -4; absolute value of standardized p-DIF > 0.1; point-biserial values < 0.2; outfit or infit values > 1.5) were flagged for review by Literably test developers. All flagged items are either discarded or retained for further analysis.

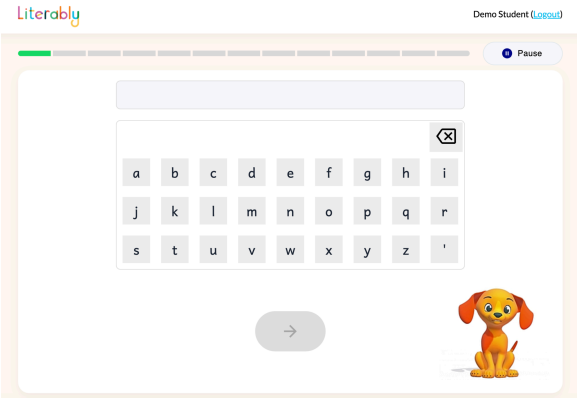
Literably developed, field-tested and validated Phonics forms for Grades K-5. However, Literably and its research partners found that Phonics did not meaningfully contribute to the

Literably Screener's predictive validity above Grade 3. Therefore, to reduce testing time, Phonics is included in the Literably Screener Composite at Grades K-3, and it is excluded at Grades 4-5.

Spelling
Grades Administered: 1-5
Construct(s) Measured: Encoding

The Literably Spelling subtest directly measures Grades 1-5 students' ability to encode phonically regular, grade-appropriate words. Literably Spelling is an untimed computer-administered assessment with standardized instructions. For each word, Literably dictates the word to the student, uses the word in a short sentence, and repeats the word. For example, "Porch. There is a swing on our front porch. Spell 'porch.'" Students spell the word using the on-screen keyboard or the keyboard on their device (e.g., iPad, computer), and the responses are scored automatically by Literably. *Figure 10* shows a sample Spelling item.

Figure 10. *Spelling - Sample Item*



To ensure developmentally-appropriate items, Literably test developers completed a literature review–including relevant research papers, standards, curricula and assessments–and developed grade-specific test blueprints that reflect the overlap of current and best practices (i.e., the skills that are taught and assessed and the skills that should be taught and assessed at each grade). The test blueprints for each grade remain the same across seasons. *Table 19* presents the skills covered for Grades 1-5.

Table 19. *Spelling - Skills Covered by Grade*

Grade	Item Types
First Grade 20 items	CVC, VCe, blends, consonant digraphs, r-controlled vowels, vowel digraphs, diphthongs, affixes, inflectional endings, 2-syllable words with closed, open, VCe, vowel team, c-le, r-controlled syllables
Second Grade 20 items	CVC,VCe, blends, consonant digraphs, r-controlled vowels, vowel digraphs, diphthongs, affixes, inflectional endings, 2- and 3-syllable words with closed, open, VCe, vowel team, c-le, r-controlled syllables

Third Grade 20 items	VCe, blends, consonant digraphs, r-controlled vowels, diphthongs, affixes, inflectional endings, 2- and 3-syllable words with closed, open, VCe, vowel team, c-le and r-controlled syllables
Fourth Grade 20 items	VCe, blends, consonant digraphs, affixes, inflectional endings, 2- and 3-syllable words with closed, open, VCe, vowel team, c-le and r-controlled syllables
Fifth Grade 20 items	VCe, r-controlled vowels, affixes, inflectional endings, 2-, 3- and 4-syllable words with closed, open, VCe, vowel team, c-le and r-controlled syllables

Once the test blueprints were set, Literably test developers wrote items aligned to the target skills. To ensure grade-appropriate items, words were selected based on age of acquisition (Kuperman et al., 2012), and each item was reviewed by multiple test developers. To avoid confusion, Literably test developers avoided homophones. Test developers also ensured no overlap in words between the Phonological Awareness, Phonics and Spelling subtests to avoid priming the student and/or over-sampling a single word.

Finally, during the 2023-2024 psychometric study, Literably and its research partners analyzed Literably Spelling item quality using classical statistics (p-value, item-total correlation) and item response theory (IRT) statistics (e.g. difficulty parameters, DIF) to ensure items were appropriately challenging, unbiased and effective at distinguishing between levels of skill on the target construct. Sub-optimal items (i.e. p-values >0.95 or <0.05; item-total correlations < 0.2; IRT difficulty > 4 or < -4; absolute value of standardized p-DIF > 0.1; point-biserial values < 0.2; outfit or infit values > 1.5) were flagged for review by Literably test developers. All flagged items are either discarded or retained for further analysis.

Oral Reading Fluency (ORF)
Grades Administered: 1-5
Construct(s) Measured: Oral Text Reading Accuracy and Rate

The Literably Oral Reading Fluency (ORF) subtest directly measures Grades 1-5 students' ability to read grade-appropriate connected text with sufficient accuracy and rate. Literably ORF is a computer-administered assessment with standardized instructions. Students read a short grade-level passage aloud. Each student's oral reading is recorded as an audio file and scored for accuracy and rate (words correct per minute). Errors, including substitutions and omissions, are marked on the graded assessment.

Literably's ORF passages were selected to resemble the materials that students read in the classroom. Thus, Literably draws its passages, with permission, from children's trade books. A team of elementary educators selected the excerpts from publishers' catalogs. Every team member had recent experience as a teacher or reading specialist serving the relevant grade(s). Literably test developers reviewed each proposed passage along the dimensions of age-appropriateness, bias, background knowledge required and syntactic and phonic difficulty. When necessary, Literably staff made small adjustments to texts to improve their conformity along these dimensions. To ensure grade-appropriateness, Literably test developers also analyzed each passage using multiple quantitative measures of text difficulty. Finally, prior to the

2023-2024 psychometric study, each ORF passage underwent extensive field testing, and passages were revised as necessary based on educator feedback.

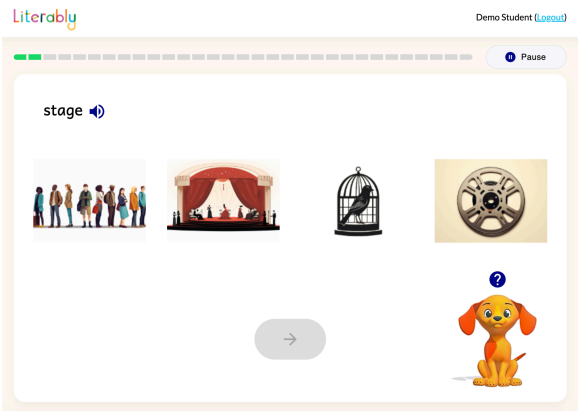
Vocabulary
Grades Administered: K-5
Construct(s) Measured: Vocabulary (receptive)

The Literably Vocabulary subtest measures Grades K-5 students' general receptive vocabulary, which strongly predicts later reading comprehension (Sénéchal et al., 2006; Scarborough, 1998; Cunningham & Stanovich, 1997). Literably Vocabulary is an untimed computer-administered assessment with standardized instructions. Literably Vocabulary includes three common and developmentally appropriate item types designed to tap different dimensions of vocabulary knowledge–picture matching items, synonym items and fill-in-the-blank items. The Kindergarten blueprint includes picture matching and synonym items, the Grades 1-2 blueprints include all three item types, and the Grades 3-5 blueprints include synonym and fill-in-the-blank items. There are a total of 20 items on each form.

Literably Vocab items feature high-frequency, high-utility words that students should be expected to know in order to comprehend texts in academic settings. Starting in Grade 1, Literably Vocab includes content-specific words across Language Arts, Math, Science and Social Studies domains. To ensure grade-appropriate items, Literably test developers selected both target words and distractors from well-established vocabulary lists (e.g. Biemiller, 2009; Marzano & Simms, 2013) based on age of acquisition (Kuperman et al., 2012), and each item was reviewed by multiple test developers.

Picture matching items
Picture matching items ask students to select the picture that best matches the target word. Students can read the word independently or have it read aloud. This picture-vocabulary format is widely-used by well-regarded assessments of receptive vocabulary, including the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007). Literably's picture matching words were selected to be concrete and visualizable. *Figure 11* shows a sample picture-matching item.

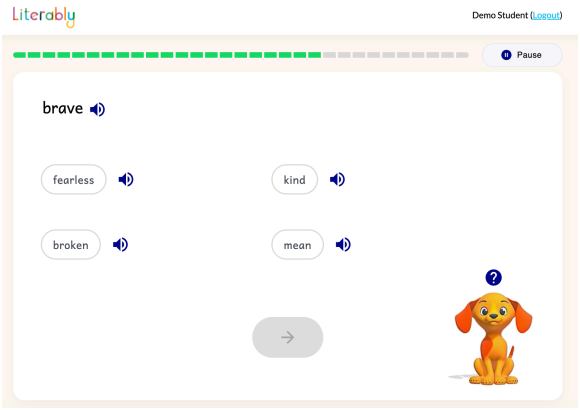
Figure 11. *Vocabulary - Sample Picture-Matching Item*



Synonym items
Synonym items ask students to select the word that is most similar in meaning to the target word. Students can read the target word and answer choices independently or have them read aloud. Synonym items require students to demonstrate knowledge of word meanings in the

absence of contextual clues. Distractors are often similar in meaning, sound and/or spelling to the target word. *Figure 12* shows a sample synonym item.

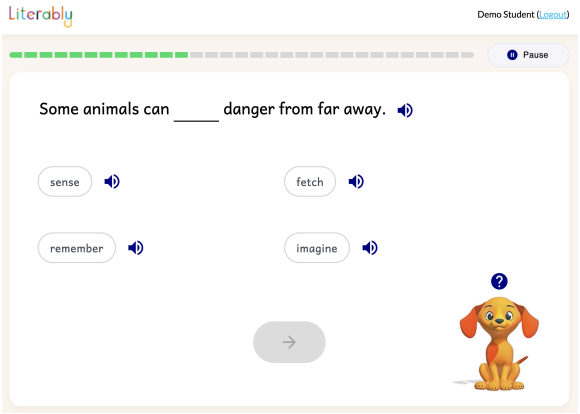
Figure 12. *Vocabulary - Sample Synonym Item*



Fill-in-the-blank items

Fill-in-the-blank items ask students to select the word that best completes the sentence. Students can read the sentence and answer choices independently or have them read aloud. Fill-in-the-blank items require students to demonstrate knowledge of word meanings in context. To ensure age-appropriate items, Grade 1 sentences are capped at 10 words, Grades 2-3 sentences are capped at 12 words, and Grades 4-5 sentences are capped at 14 words. Answer blanks typically appear near the end of the sentence to minimize the need to reread the sentence. Distractors are often similar in meaning, sound and/or spelling to the target word. *Figure 13* shows a sample fill-in-the-blank item.

Figure XX. *Vocabulary - Sample Fill-in-the-Blank Item*



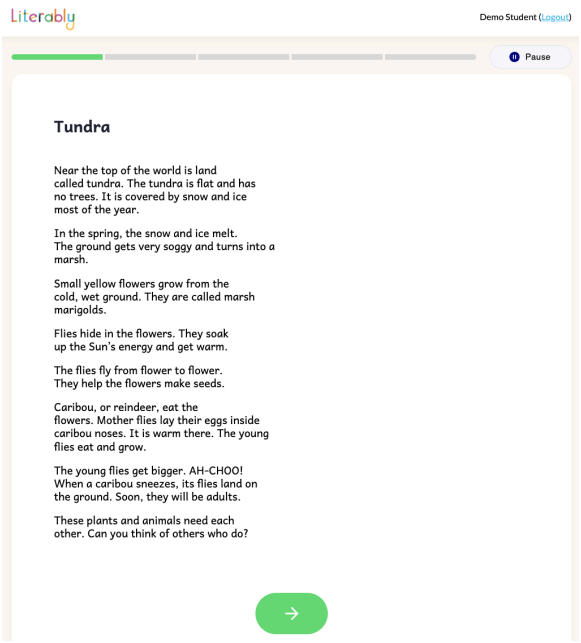
Finally, during the 2023-2024 psychometric study, Literably and its research partners analyzed Literably Vocab item quality using classical statistics (p-value, item-total correlation) and item response theory (IRT) statistics (e.g. difficulty parameters, DIF) to ensure items were appropriately challenging, unbiased and effective at distinguishing between levels of skill on the target construct. Sub-optimal items (i.e. p-values >0.95 or <0.05; item-total correlations < 0.2; IRT difficulty > 4 or < -4; absolute value of standardized p-DIF > 0.1; point-biserial values < 0.2; outfit or infit values > 1.5) were flagged for review by Literably test developers. All flagged items are either discarded or retained for further analysis.

Comprehension
Grades Administered: 2-5
Construct(s) Measured: Reading Comprehension

The Literably Comprehension subtest measures Grades 2-5 students' ability to read and understand grade-appropriate fiction and nonfiction texts. Literably Comprehension is an untimed computer-administered assessment with standardized instructions. The format of Literably Comprehension mirrors authentic passage-and-question tasks that students often complete in the classroom. Each Literably Comprehension assessment includes two passages - one fiction and one nonfiction. Each passage is followed by six multiple-choice comprehension questions, presented one at a time. Students can refer to the passage while answering the questions, and they can return to previous questions from the same passage. However, once the student has progressed to the second passage, they cannot return to the first passage.

Literably Comprehension passages were excerpted from authentic children's literature and nonfiction works. Literably test developers selected the excerpts and reviewed each passage along the dimensions of age-appropriateness, bias, background knowledge required and syntactic and phonic difficulty. When necessary, Literably staff made small adjustments to texts to improve their conformity along these dimensions. Literably test developers used Flesch-Kincaid readability estimates to determine the grade-appropriateness of the passages. *Figure 14* shows a sample passage.

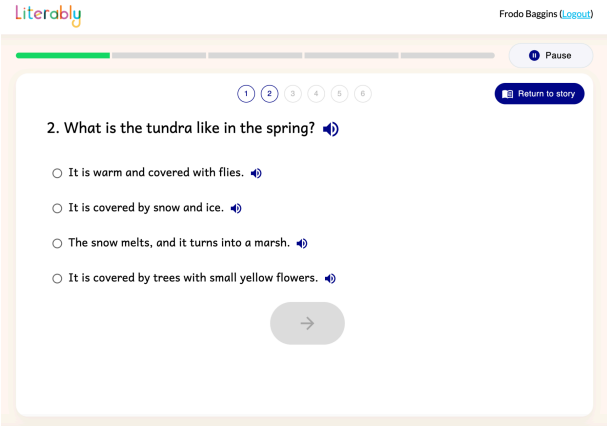
Figure 14. *Comprehension - Sample Passage*



Literably test developers worked with a team of experienced educators to write the comprehension questions and then revised them based on feedback from a broader pool of educators. To maximize alignment with the curricula of schools in the United States, the authors wrote every comprehension question to assess one or more of the Common Core State Standards (Common Core State Standards Initiative, 2010). Literably Comprehension includes

both literal and inferential questions. Questions were written to be highly text-dependent, such that they would be difficult to answer correctly without an understanding of the text. *Figure 15* shows a sample question.

Figure 15. *Comprehension - Sample Question*



Finally, during the 2023-2024 psychometric study, Literably and its research partners analyzed Literably Comprehension item quality using classical statistics (p-value, item-total correlation) and item response theory (IRT) statistics (e.g. difficulty parameters, DIF) to ensure items were appropriately challenging, unbiased and effective at distinguishing between levels of skill on the target construct. Sub-optimal items (i.e. p-values >0.95 or <0.05; item-total correlations < 0.2; IRT difficulty > 4 or < -4; absolute value of standardized p-DIF > 0.1; point-biserial values < 0.2; outfit or infit values > 1.5) were flagged for review by Literably test developers. All flagged items are either discarded or retained for further analysis.

Construct Validity

External Evidence: Correlations to Other Measures

Literably researchers evaluated construct validity by correlating each Literably Screener subtest with each DIBELS 8 subtest (at Grades K-2) and with the SBAC ELA score (at Grades 3-5).

Tables 21-24 present the correlation matrices for Grades K-5. To describe correlations, we've adopted the classifications from Hopkins (2022) shown in *Table 20*. As expected, most correlations among subtests that measure similar constructs are moderate to strong. Correlations are generally weaker among subtests that measure distinct constructs.

Table 20. *Correlation Coefficients Descriptions*

Coefficient Range	Descriptor
.70 - 1	Strong
.50 – .69	Moderate-Strong
.30 – .49	Moderate
.10 – .29	Small

0 - .09	Very Small

Table 21. *Kindergarten correlations between Literably subtests (Winter) and DIBELS 8 subtests (Spring)*

N=53	DIBELS LNF	DIBELS PSF	DIBELS NWF	DIBELS WRF
Literably RAN	0.44	0.24	0.26	0.46
Literably PA	0.29	0.23	0.37	0.51
Literably Phonics	0.44	0.22	0.35	0.63
Literably Vocab	0.27	0.10	0.15	0.36

Table 22. *Grade 1 correlations between Literably subtests (Winter) and DIBELS 8 subtests (Spring)*

N=112	DIBELS LNF	DIBELS PSF	DIBELS NWF	DIBELS WRF	DIBELS ORF
Literably RAN	0.43	0.36	0.40	0.48	0.44
Literably PA	0.34	0.53	0.47	0.50	0.49
Literably Phonics	0.19	0.26	0.43	0.53	0.49
Literably Spelling	0.38	0.30	0.54	0.63	0.63
Literably ORF	0.38	0.25	0.57	0.73	0.74
Literably Vocab	0.19	0.25	0.37	0.45	0.48

Table 23. *Grade 2 correlations between Literably subtests (Winter) and DIBELS 8 subtests (Spring)*

N=99	DIBELS NWF	DIBELS WRF	DIBELS ORF	DIBELS Maze
Literably Phonics	0.35	0.45	0.57	0.49
Literably Spelling	0.46	0.59	0.78	0.68
Literably ORF	0.42	0.65	0.88	0.72

Literably Vocab	0.30	0.46	0.57	0.44
Literably Comprehension	0.28	0.42	0.57	0.54

Table 24. *Grades 3-5 correlations between Literably subtests (Winter) and SBAC ELA (Spring)*

Literably Subtest	Grade 3 (N=165)	Grade 4 (N=153)	Grade 5 (N=110)
Phonics	0.42	-	-
Spelling	0.62	0.69	0.60
ORF	0.56	0.49	0.62
Vocab	0.59	0.74	0.70
Comprehension	0.62	0.54	0.61

Internal Evidence: Correlations among Literably Subtests

Literably and its research partners further evaluated construct validity by correlating the Literably subtests to each other.

Tables 25-30 present the correlation matrices for Grades K-5. As expected, the correlations are stronger among subtests that measure similar constructs, and weaker among subtests that measure distinct constructs. Most correlations among similar subtests are moderate to strong.

Table 25. *Kindergarten correlations among Literably subtests (Fall)*

N=86	RAN	PA	Phonics	Vocab
RAN	-	0.58	0.53	0.48
PA	0.58	-	0.74	0.54
Phonics	0.53	0.74	-	0.53
Vocab	0.48	0.54	0.53	-

Table 26. *Grade 1 correlations among Literably subtests (Fall)*

N=94	RAN	PA	Phonics	Spelling	ORF	Vocab
RAN	-	0.27	0.38	0.32	0.43	0.22
PA	0.27	-	0.61	0.47	0.33	0.53
Phonics	0.38	0.61	-	0.69	0.60	0.52
Spelling	0.32	0.47	0.69	-	0.55	0.51

ORF	0.43	0.33	0.60	0.55	-	0.48
Vocab	0.22	0.53	0.52	0.51	0.48	-

Table 27. *Grade 2 correlations among Literably subtests (Fall)*

N=93	Phonics	Spelling	ORF	Vocab	Comprehension
Phonics	-	0.60	0.63	0.54	0.46
Spelling	0.60	-	0.70	0.52	0.58
ORF	0.63	0.70	-	0.45	0.51
Vocab	0.54	0.52	0.45	-	0.54
Comprehension	0.46	0.58	0.51	0.54	-

Table 28. *Grade 3 correlations among Literably subtests (Fall)*

N=161	Phonics	Spelling	ORF	Vocab	Comprehension
Phonics	-	0.51	0.50	0.39	0.48
Spelling	0.51	-	0.64	0.52	0.44
ORF	0.50	0.64	-	0.53	0.50
Vocab	0.39	0.52	0.53	-	0.52
Comprehension	0.48	0.44	0.50	0.52	-

Table 29. *Grade 4 correlations among Literably subtests (Fall)*

N=189	Spelling	ORF	Vocab	Comprehension
Spelling	-	0.67	0.57	0.55
ORF	0.67	-	0.55	0.47
Vocab	0.57	0.55	-	0.66
Comprehension	0.55	0.47	0.66	-

Table 30. *Grade 5 correlations among Literably subtests (Fall)*

N=150	Spelling	ORF	Vocab	Comprehension

Spelling	-	0.72	0.57	0.56
ORF	0.72	-	0.53	0.53
Vocab	0.57	0.53	-	0.57
Comprehension	0.56	0.53	0.57	-

Taken together, these results - internal and external, convergent and discriminant - provide evidence that the Literably Screener subtests effectively measure the intended constructs.

Criterion Validity

For criterion validity, Literably and its research partners examined the relationship between the Literably Screener Composite administered in the Winter of the 2023-2024 school year and DIBELS 8 and Smarter Balanced Assessment Consortium (SBAC) criterion measures collected in Spring 2024.

The Literably Screener Composite Score

The Literably Screener Composite Score is a simple unweighted sum of scores from select Literably measures designed to provide an accurate and efficient estimate of a student's overall reading proficiency.

To develop the Literably Screener Composite Score, Literably and its research partners performed a series of analyses to determine which Literably measures, for each grade, were most predictive of end-of-year reading proficiency. *Table 31* reports the measures that contribute to the Literably Screener Composite Score for Grades K-5.

Table 31. *Literably Screener Composite measures by grade*

Grade	*Literably Screener Composite Measures*
K	RAN, PA, Phonics, Vocabulary
1	RAN, PA, Phonics, Spelling, ORF, Vocabulary
2	Phonics, Spelling, ORF, Vocabulary, Comprehension
3	Phonics, Spelling, ORF, Vocabulary, Comprehension
4	Spelling, ORF, Vocabulary, Comprehension
5	Spelling, ORF, Vocabulary, Comprehension

Note: Literably offers additional measures beyond those that contribute to the Composite, and teachers can use those measures to gain additional insights to inform their instruction.

The Criterion Measures

At Grades K-2, Literably reports criterion validity relative to measures from DIBELS 8. At Grades 3-5, Literably reports criterion validity relative to the Smarter Balanced Assessment Consortium (SBAC) ELA score. *Table 32* shows the selected criterion measures for grades K-5.

Table 32. *Criterion Measures by Grade*

Grade	*Criterion*
K	DIBELS 8 Word Reading Fluency
1	DIBELS 8 Oral Reading Fluency - Words Read Correctly
2	DIBELS 8 Composite, including Nonsense Word Fluency, Word Reading Fluency, Oral Reading Fluency and Maze
3	SBAC ELA
4	SBAC ELA
5	SBAC ELA

DIBELS 8

DIBELS 8 is widely-used and well-researched set of teacher-administered early literacy assessments that have demonstrated strong correlations with broad measures of reading proficiency (e.g. Iowa Total Reading), and many literacy screening assessments have validated against measures from the DIBELS/Acadience family.

At each grade level, Literably researchers selected as the criterion the DIBELS 8 score that best represents the desired outcome of instruction at that grade, as described here:

- At Kindergarten, the DIBELS 8 subtests are Letter Naming Fluency (LNF), Phonemic Segmentation Fluency (PSF), Nonsense Word Fluency (NWF) and Word Reading Fluency (WRF). In the judgment of Literably researchers, Word Reading Fluency is the most appropriate criterion, because it best represents the goal of kindergarten reading instruction (fluent reading of real words), whereas LNF, PSF and NWF are better understood as predictors of future performance. WRF is also among the DIBELS 8 kindergarten measures that correlates most strongly with Iowa Total Reading (even more strongly than the DIBELS 8 Composite).
- At Grade 1, the DIBELS 8 subtests are LNF, PSF, NWF, WRF and Oral Reading Fluency (ORF). Of these, Literably researchers determined that ORF is the most appropriate criterion, because it best represents the goal of Grade 1 instruction (fluent reading of connected text). ORF is also the DIBELS 8 Grade 1 measure that correlates most strongly with Iowa Total Reading (even more strongly than the DIBELS 8 Composite).
- At Grade 2, the DIBELS 8 subtests are NWF, WRF, ORF and Maze (a measure of reading comprehension). Literably researchers determined that the most appropriate Grade 2 criterion is the DIBELS 8 Composite Score, because it best represents the goal of Grade 2 instruction (fluent reading with comprehension). The DIBELS 8 Composite Score is also the DIBELS 8 Grade 1 score that correlates most strongly with Iowa Total Reading.

SBAC

The Smarter Balanced Assessment Consortium SBAC ELA summative assessment is a well-validated, standardized, and standards-aligned measure of English Language Arts proficiency that is used as the state test in 18 states, serving grades 3-8 and 10, and many literacy screening assessments have validated against SBAC ELA or similar state tests.

Predictive Validity

Table 33 presents predictive validity for the Literably Screener Composite relative to the selected criterion measures administered in Spring 2024. All of the Literably Winter administrations were completed more than 2 months before the Spring criterion measures.

Table 33. *Predictive Validity for the Literably Screener Composite*

Grade	*Criterion*	N	r
K	WRF	53	0.55
1	ORF	112	0.70
2	Composite	99	0.67
3	SBAC	178	0.66
4	SBAC	153	0.61
5	SBAC	122	0.69

To describe correlations, we've adopted the classifications from Hopkins (2022) shown in *Table 34*.

Table 34. *Correlation Coefficients Descriptions*

Coefficient Range	*Descriptor*
.70 - 1	Strong
.50 – .69	Moderate-Strong
.30 – .49	Moderate
.10 – .29	Small
0 - .09	Very Small

All correlations between the Literably Screener Composite and the selected criterion measures are moderate-strong to strong and all exceed 0.5.

These predictive correlations with DIBELS 8 and SBAC criterion measures reflect well on the Literably Screener as a valid measure of reading performance and as a predictor of end-of-year reading proficiency at Grades K-5.

Classification Accuracy

Definition of Risk

Literably and its research partners evaluated the Literably Screener's classification accuracy relative to student performance on selected DIBELS 8 and SBAC criterion measures administered in Spring 2024. The appropriateness of DIBELS 8 and SBAC as criterion measures is discussed above under criterion validity.

Literably researchers considered a student to be "high risk" for reading difficulties if they scored below the 20th percentile on the end-of-year criterion measure, which conforms to the National Center on Intensive Intervention (NCII) guidelines for definitions of risk.

Table 35 shows how Literably researchers defined risk based on criterion measure performance at grades K-5.

Table 35. *Risk Classifications based on Criterion Measure Performance*

Grade	*High Risk*	*Criterion Measure*
K	<20th percentile	DIBELS 8 Word Reading Fluency
1	<20th percentile	DIBELS 8 Oral Reading Fluency - Words Read Correctly
2	<20th percentile	DIBELS 8 Composite, including Nonsense Word Fluency, Word Reading Fluency, Oral Reading Fluency and Maze
3	<20th percentile	SBAC ELA
4	<20th percentile	SBAC ELA
5	<20th percentile	SBAC ELA

Cut Score Methodology

Literably researchers established cut scores for the Literably Composite and for each Literably subtest for each grade and season.

For each score, if some of the potential cut scores yielded sensitivity and specificity above 0.8, then Literably researchers selected the cut score that maximized sensitivity while maintaining specificity above 0.8. If some of the potential cut scores yielded sensitivity and specificity above 0.7 (but not 0.8), then Literably researchers selected the cut score that maximized sensitivity while maintaining specificity over 0.7. Finally, if none of the potential cut scores yielded sensitivity and specificity above 0.7, then Literably researchers selected the cut score that maximized the sum of sensitivity and specificity.

Composite Cut Scores and Classification Accuracy Estimates

The Literably Composite cut scores can be used to identify students who are at high risk for reading difficulties (below the 20th percentile). Scores below the cut score are color-coded red, and these students are in need of intensive support.

Table 36 reports the Literably Screener's cut score, area under the curve (AUC), sensitivity and specificity by grade and season.

***Table 36**. Literably Composite Sensitivity, Specificity and AUC by Grade, Season and Cut*

Grade	Season	N	Cut Score	Sensitivity	Specificity	AUC
K	Fall	86	39.07	0.86	0.82	0.91
	Winter	53	42.45	1	0.96	0.98
	Spring	65	70.38	1	0.93	0.94
1	Fall	94	93.41	1	0.91	0.97
	Winter	112	95.72	1	0.90	0.98
	Spring	129	112.94	1	0.97	0.98
2	Fall	93	49	1	0.95	0.99
	Winter	99	59	1	0.91	0.99
	Spring	109	72	1	0.92	0.99
3	Fall	161	118	0.81	0.74	0.85
	Winter	178	103	0.82	0.71	0.84
	Spring	174	118	0.82	0.79	0.85
4	Fall	189	114	0.91	0.83	0.91
	Winter	153	128	0.88	0.76	0.86
	Spring	161	118	0.88	0.85	0.89
5	Fall	150	121	0.80	0.68	0.78
	Winter	122	119	0.80	0.78	0.82
	Spring	125	120	0.88	0.80	0.90

NCII rates a screening tool highest when AUC, sensitivity and specificity estimates meet or exceed 0.8. AUC, sensitivity and specificity estimates between 0.8 and 0.7 are generally considered acceptable.

For the Literably Screener Composite, all sensitivity estimates exceed 0.8, and half exceed 0.9. All but one specificity estimate exceeds 0.7, most exceed 0.8, and many exceed 0.9. All but one AUC estimate exceeds 0.8 and most exceed 0.9. These results suggest that the Literably Screener is very effective at identifying students at high risk of reading difficulties.

Note: In addition to the primary Composite cut score, Literably researchers are currently developing supplementary cut scores that predict moderate and low risk. These additional cut scores will be available in advance of the 2025-2026 school year.

Subtest Cut Scores and Classification Accuracy Estimates

In addition to the Literably Composite cut scores, Literably researchers established cut scores for each subtest to help guide instructional decision-making. The Literably subtest cut scores can be used to identify students who - based on their performance on a specific subtest - appear to be at high risk for reading difficulties (below the 20th percentile). Scores below the cut score are color-coded red, and these students are in need of intensive support.

Table 37 reports the cut score, area under the curve (AUC), sensitivity and specificity for each Literably subtest, grade and season.

***Table 37**. Literably Subtests' Sensitivity, Specificity and AUC by Grade and Season*

Grade	Season (N)	Subtest	Cut Score	Sensitivity	Specificity	AUC
K	Fall (86)	RAN	18.7	0.86	0.89	0.91
		PA	3	0.71	0.85	0.84
		Phonics	5	0.86	0.73	0.83
		Vocab	7	0.71	0.67	0.67
	Winter (53)	RAN	23.2	1	0.96	0.96
		PA	4	1	0.85	1
		Phonics	4	1	0.81	0.95
		Vocab	11	1	0.71	0.73
	Spring (65)	RAN	40.6	0.80	0.93	0.91
		PA	10	1	0.60	0.82
		Phonics	7	0.80	0.82	0.86
		Vocab	11	0.60	0.75	0.85
1	Fall (94)	RAN	53.4	1	0.80	0.91
		PA	4	0.67	0.95	0.79
		Phonics	7	1	0.82	0.91
		Spelling	3	0.67	0.77	0.75
		ORF	11	1	0.79	0.91

		Vocab	6	0.67	0.85	0.66
	Winter (112)	RAN	52.6	1	0.91	0.96
		PA	6	0.83	0.82	0.89
		Phonics	8	0.67	0.70	0.80
		Spelling	5	0.83	0.73	0.86
		ORF	13	0.83	0.83	0.94
		Vocab	9	0.83	0.72	0.82
	Spring (129)	RAN	62.1	1	0.87	0.93
		PA	6	1	0.83	0.89
		Phonics	8	1	0.87	0.92
		Spelling	5	1	0.76	0.84
		ORF	51	1	0.65	0.82
		Vocab	9	1	0.82	0.86
2	Fall (93)	Phonics	5	0.90	0.88	0.94
		Spelling	3	0.90	0.82	0.93
		ORF	22	1	0.95	0.98
		Vocab	12	0.70	0.66	0.76
		Compr	3	0.70	0.84	0.81
	Winter (99)	Phonics	5	1	0.90	0.97
		Spelling	3	1	0.83	0.93
		ORF	37	1	0.88	0.97
		Vocab	12	0.78	0.78	0.85
		Compr	4	0.89	0.66	0.78
	Spring (109)	Phonics	7	0.90	0.84	0.88
		Spelling	4	1	0.85	0.94
		ORF	40	1	0.94	0.99
		Vocab	12	0.80	0.75	0.85

		Compr	5	1	0.71	0.88
3	Fall (161)	Phonics	14	0.78	0.64	0.79
		Spelling	4	0.75	0.75	0.83
		ORF	76	0.75	0.80	0.83
		Vocab	10	0.63	0.74	0.70
		Compr	5	0.73	0.82	0.73
	Winter (178)	Phonics	14	0.85	0.68	0.80
		Spelling	5	0.79	0.74	0.83
		ORF	67	0.72	0.70	0.79
		Vocab	11	0.74	0.73	0.79
		Compr	6	0.63	0.81	0.81
	Spring (174)	Phonics	16	0.82	0.72	0.81
		Spelling	6	0.79	0.71	0.84
		ORF	82	0.85	0.72	0.80
		Vocab	11	0.73	0.72	0.81
		Compr	4	0.92	0.64	0.74
4	Fall (189)	Spelling	7	0.76	0.78	0.84
		ORF	95	0.88	0.71	0.87
		Vocab	11	0.82	0.82	0.89
		Compr	5	0.73	0.82	0.87
	Winter (153)	Spelling	7	0.83	0.84	0.90
		ORF	98	0.79	0.73	0.79
		Vocab	11	0.79	0.85	0.90
		Compr	6	0.63	0.81	0.77
	Spring (161)	Spelling	8	0.83	0.80	0.86
		ORF	99	0.71	0.70	0.82
		Vocab	12	0.79	0.80	0.86

		Compr	4	0.92	0.64	0.79
5	Fall (150)	Spelling	7	0.70	0.75	0.80
		ORF	97	0.70	0.64	0.74
		Vocab	11	0.80	0.75	0.89
		Compr	5	0.73	0.78	0.87
	Winter (122)	Spelling	9	0.80	0.71	0.82
		ORF	97	0.72	0.75	0.78
		Vocab	11	0.76	0.73	0.85
		Compr	5	0.80	0.58	0.77
	Spring (125)	Spelling	10	0.80	0.73	0.86
		ORF	86	0.80	0.84	0.87
		Vocab	11	0.84	0.80	0.90
		Compr	3	0.76	0.79	0.83

NCII rates a screening tool highest when AUC, sensitivity and specificity estimates meet or exceed 0.8. AUC, sensitivity and specificity estimates between 0.8 and 0.7 are generally considered acceptable.

It is generally not expected that each subtest would individually meet NCII's criteria. Nevertheless, of the 252 sensitivity, specificity and AUC estimates above, over 60% exceed 0.8, and over 90% exceed 0.7. These results suggest that the Literably Screener subtests are effective at identifying students at high risk of reading difficulties.

Note: In addition to the high risk subtest cut scores presented above, Literably researchers are currently developing supplementary subtest cut scores that predict moderate and low risk. These additional cut scores will be available in advance of the 2025-2026 school year.

Bias Analyses

To ensure the Literably Screener does not exhibit bias towards any major subgroups, Literably researchers analyzed differential item functioning for the item-based subtests and validity by subgroup for the composite score.

Differential Item Functioning

For Literably's item-based assessments (i.e. Phonological Awareness, Phonics, Spelling, Vocab, Comprehension), to ensure items are not biased against subgroups, Literably and its research partners evaluated differential item functioning (DIF) using standardized P-DIF. The sample sizes were sufficient to examine DIF in relation to Gender, Ethnicity, English Learner

status, Special Ed Status, and FRPL-eligibility. The comparisons examined were Female vs. Male; Black, Hispanic and Asian vs. White; English Learner vs. non-English Learner, Special Ed vs. General Ed, and FRPL-eligible vs. FRPL-ineligible.

The DIF results are described using the categories presented in *Table 38*.

Table 38. *DIF Categories*

Category	*p-DIF Absolute Value*
Negligible	< 0.1
Moderate	0.1 - 0.25
Severe	>0.25

Across all subtests and comparisons, the vast majority of items demonstrated negligible DIF. Sub-optimal items (p-DIF absolute value > 0.1) were flagged for review by Literably test developers, and all flagged items were either discarded or retained for further analysis.

Composite Validity by Subgroup

Table 39 presents predictive validity for the Literably Screener Composite Score by subgroup. All coefficients are moderate to strong, providing evidence that the Literably Screener is a valid measure of literacy across subgroups.

Table 39. *Predictive Validity by Subgroup for the Literably Screener Composite (Winter)*

Grade	*Criterion*	*Subgroup*	N	r
K	DIBELS WRF	Male	27	0.58
		Female	25	0.47
		White	41	0.50
		FRPL eligible	32	0.51
		SPED	9	0.34
1	DIBELS ORF	Male	41	0.67
		Female	46	0.62
		White	43	0.64
		Hispanic	19	0.75
		Asian	15	0.69
		FRPL eligible	47	0.55
2	DIBELS	Male	46	0.58

	Composite	Female	44	0.78
		White	53	0.67
		Hispanic	17	0.91
		FRPL eligible	52	0.74
		English Learner	10	0.91
		Sp. Ed.	14	0.66
3	SBAC ELA	Male	85	0.69
		Female	93	0.63
		White	90	0.66
		Hispanic	50	0.58
		Asian	13	0.79
		FRPL eligible	106	0.60
		English Learner	29	0.60
		Sp. Ed.	13	0.81
4	SBAC ELA	Male	71	0.56
		Female	82	0.68
		White	73	0.70
		Hispanic	28	0.49
		Asian	30	0.47
		Multiracial	11	0.60
		FRPL eligible	90	0.58
		English Learner	16	0.42
		Sp. Ed.	13	0.59
5	SBAC ELA	Male	64	0.65
		Female	58	0.72
		White	53	0.63
		Hispanic	15	0.72

		Black	11	0.80
		Asian	23	0.72
		Multiracial	14	0.76
		FRPL eligible	83	0.67
		English Learner	19	0.74
		Sp. Ed.	12	0.56

Item Difficulty

To ensure the Literably Screener includes items of varied difficulty, Literably researchers calculated Item response theory (IRT) item difficulty parameters for the item-based assessments (i.e. PA, Phonics, Spelling, Vocab, Comprehension). The Literably Screener item pool includes items that are easy, moderate and challenging. *Table 40* below shows the distribution of item difficulty parameters. These results provide evidence that the Literably Screener includes items across an appropriately broad range of difficulty.

Table 40. *Distribution of Item Difficulty Parameters*

Item Difficulty	Percentage
< -1.5	9%
-1.5 to 1.5	81%
> 1.5	9%

Interpretation of Scores

Overview

The Response to Intervention (RTI) model and the multi-tiered system of support (MTSS) framework that encompasses it have been implemented widely in U.S. schools to guide data-based decision making in the past two decades. The RTI model involves screening to identify students at risk, a multi-level prevention system, progress monitoring, and evidence-based instruction and interventions (NCRTI, 2016).

The Literably Screener provides data that schools can use to support students along the MTSS continuum. As described below, Literably's validated performance classifications allow school systems to identify students who should participate in three different tiers of support, with recommendations for best practices at each tier.

Composite Score Performance Classifications

After students complete the Literably Screener during each testing period, Literably will report individual raw scores for each subtest and an overall Composite Score for the screener. The Composite Score provides the best overall estimate of students' early literacy skills and likelihood of reading success.

For easy interpretation, the Literably Screener composite score is translated into performance classifications and color-coded based on empirically-established Literably cut scores. The 4 performance classifications are exceeds (blue), meets (green), approaches (yellow), and below (red).

This system of performance classifications allows educators to identify students' levels of risk and informs decisions related to appropriate interventions. It is recommended that school systems analyze data from the Literably Screener in the context of other classroom data and collaboratively (e.g., within a grade level team and with the involvement of learning specialists and parents) in order to design interventions within an MTSS framework. For ELs, classifications and their interpretations should be reviewed in the context of other available data, such as the results of screening in the student's home language, if available, and the student's educational history.

School systems can use the following guidance for interpreting Literably screening classifications and determining instructional next steps.

Performance Classification	**Interpretation and Next Steps**
Exceeds (blue)	Interpretation: A student classified as "exceeds" is very likely to achieve grade-level literacy goals. Next Steps: The student should be well-served by high-quality, research-based core reading instruction (Tier 1).
Meets (green)	Interpretation: A student classified as "meets" is likely to achieve grade-level literacy goals. Next Steps: The student should be well-served by high-quality, research-based core reading instruction (Tier 1), but individual monitoring and specific support may be appropriate for some students in this category, especially those who score on the lower end of the score range.
Approaches (yellow)	Interpretation: A student classified as "approaches" is at some risk of failing to achieve grade-level literacy goals. Next Steps: The student would benefit from supplemental support in specific skills in addition to effective core instruction. The student should receive targeted small-group intervention and progress monitoring (Tier 2). This may entail weekly small group instruction by a learning specialist in some of the foundational reading skills, using an evidence-based, structured, and explicit approach to instruction. Intervention should be accompanied by progress monitoring to ensure the student makes adequate progress.

Below (red)	Interpretation: A student classified as below is performing well-below grade level expectations, and is at risk of severe difficulties. Next Steps: The student would benefit from intensive support (Tier 3) in addition to effective core instruction. ● Intensive interventions for students in this category may include the teaching of prerequisite (i.e., below grade-level) skills, delivering instruction in a small group or individually, providing more instructional time or practice, or providing more explicit modeling and scaffolding. ● The student should be monitored routinely. If the student does not make adequate progress after a period of interventions, it may be appropriate to refer the student for additional diagnostic testing to identify specific skill deficits and analyze their underlying causes, with the goal of designing more effective interventions and supports.

Note: This Technical Manual presents evidence related to the low cut score that identifies students at high risk of reading difficulties (red). Literably researchers have already developed the medium and high cut scores for the K-2 Composite (which will distinguish yellow, green and blue) and are currently developing the medium and high cut scores for the 3-5 Composite and K-5 subtests. These additional cut scores will be available in advance of the 2025-2026 school year.

References

Anthony, J.L., & Francis, D.J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14*(5), 255–259.

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology, 107*(3), 868–883. https://doi.org/10.1037/edu0000006

Biemiller, A. (2009). Words Worth Teaching. New York: SRA.

Catts HW, Fey ME, Zhang X, Tomblin JB. Estimating the Risk of Future Reading Difficulties in Kindergarten Children: A Research-Based Model and Its Clinical Implementation. Lang Speech Hear Serv Sch. 2001 Jan 1;32(1):38-50. doi: 10.1044/0161-1461(2001/004). PMID: 27764435.

Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading, 6*(4), 369–400. https://doi.org/10.1207/S1532799XSSR0604_04

Clemens, N. H., Shapiro, E. S., Wu, J.-Y., Taylor, A. B., & Caskie, G. L. (2014). Monitoring early first-grade reading progress: A comparison of two measures. *Journal of Learning Disabilities, 47*(3), 254–270. https://doi.org/10.1177/0022219412454455

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934–945. https://doi.org/10.1037/0012-1649.33.6.934

Dale, E., & O'Rourke, J. (1981). The Living Word Vocabulary: A National Vocabulary Inventory. Chicago: World Book.

Denckla, M. B., & Rudel, R. G. (1976). Rapid "automatized" naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14*(4), 471–479. https://doi.org/10.1016/0028-3932(76)90075-0

Deno, S. L., Mirkin, P. K., Chiang, B., and Lowry, L. (1980). Relationships Among Simple Measures of Reading and Performance on Standardized Achievement Tests. (Research Report No. 20). Minneapolis, MN: Institute for Research on Learning Disabilities, University of Minnesota.

Dunn, L. M., & Dunn, D. M. (2007). Peabody picture vocabulary TestIV. Circle Pines, MN: American Guidance Service.

Ehri LC, Nunes SR, Willows DM, Schuster BV, Yaghoub-Zadeh Z, Shanahan T. Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly.* 2001;36:250–287.

Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, DHHS. (2000). Report of the National Reading Panel: Teaching Children to Read: Reports of the Subgroups (00-4754). Washington, DC: U.S. Government Printing Office.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). Macmillan.

Fletcher JM, Francis DJ, Foorman BR, Schatschneider C. Early Detection of Dyslexia Risk: Development of Brief, Teacher-Administered Screens. Learn Disabil Q. 2021 Aug;44(3):145-157. doi: 10.1177/0731948720931870. Epub 2020 Jun 19. PMID: 34584341; PMCID: PMC8475291.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring Early Reading Development in First Grade: Word Identification Fluency Versus Nonsense Word Fluency. *Exceptional Children, 71*(1), 7–21. https://doi.org/10.1177/001440290407100101

Georgiou, George & Parrila, Rauno & Manolitsis, George & Kirby, John. (2011). Examining the importance of assessing rapid automatized naming (RAN) speed for the identification of children with reading difficulties. Learning Disabilities. 9. 5-26.

Good, R. H., & Kaminski, R. A. (Eds.) (2002). Dynamic Indicators of Basic Early Literacy Skills (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: http://dibels.uoregon.edu.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *RASE: Remedial & Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Hjetland, H. N., Brinchmann, E. I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: A systematic review. *Campbell Systematic Reviews*, *13*(1), 1–155. https://doi.org/10.4073/csr.2017.14

Hosp, M. K., & Hosp, J. (2003). Curriculum-based measurement for reading, math, and spelling: How to do it and why. Preventing School Failure, 48(1), 10–17.

Howe, K. B., & Shinn, M. M. (2002). Standard Reading Assessment Passages (RAPs) for use in General Outcome Measurement: A manual describing development and technical features. Eden Prairie, MN: Edformation, Inc.

Katzir, Tami & Kim, Young-Suk & Wolf, Maryanne & O'Brien, Beth & Kennedy, Becky & Lovett, Maureen & Morris, Robin. (2006). Reading Fluency: The whole is more than the parts. Annals of Dyslexia. 56. 51-82. 10.1007/s11881-006-0003-5.

Kilgus, S. P., Methe, S. A., Maggin, D. M., & Tomasula, J. L. (2014). Curriculum-based measurement of oral reading (R-CBM): A diagnostic test accuracy meta-analysis of evidence supporting use in universal screening. *Journal of School Psychology, 52*(4), 377–405. https://doi.org/10.1016/j.jsp.2014.06.002

Kilpatrick, D. A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties.* Hoboken, NJ: Wiley.

Kuperman, V., Stadthagen-Gonzalez, H. & Brysbaert, M. Age-of-acquisition ratings for 30,000 English words. *Behav Res* 44, 978–990 (2012). https://doi.org/10.3758/s13428-012-0210-4

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science, 20,* 1040–1048. doi:10.1111/j.1467- 9280.2009.02405.x

Lesaux, N. K., & Siegel, L. S. (2003). The Development of Reading in Children Who Speak English as a Second Language. *Developmental Psychology, 39*(6), 1005–1019. https://doi.org/10.1037/0012-1649.39.6.1005

Lonigan CJ, Burgess SR, Anthony JL. Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology.* 2000;36:596–613.

Marzano, R.J., & Simms, J.A. (2013). Vocabulary for the Common Core. Centennial, CO: Marzano Research.

McWeeny, S., Choi, J. S., Choe, J., LaTourette, A., Roberts, M. Y., & Norton, E. S. (2022). Rapid automatized naming (RAN) as a kindergarten predictor of future reading in English: A systematic review and meta-analysis. *Reading Research Quarterly, 57*(4), 1187–1211. DOI: 10.1002/rrq.467

Moats, Louisa. (2005). How Spelling Supports Reading And Why It Is More Regular and Predictable Than You May Think. American Educator. 29.

National Center on Response to Intervention (NCRTI). (2016). Essential Components of MTSS. Retrieved from https://mtss4success.org/essential-components

National Center on Response to Intervention (January 2013). Screening Briefs Series—Brief #3: Predicting Students at Risk for Reading and Mathematics Difficulties. Washington, DC: U.S.

Department of Education, Office of Special Education Programs, National Center on Response to Intervention.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: Authors.

National Research Council. 1998. *Preventing Reading Difficulties in Young Children*. Washington, DC: The National Academies Press. https://doi.org/10.17226/6023.

Price, P., Tepperman, J., Iseli, M., Duong, T., Black, M., Wang, S., ... & Alwan, A. (2009). Assessment of emerging reading skills in young native speakers and language learners. Speech Communication, 51 (10), 968984.

Rathvon, N. (2004). Early reading assessment: A practitioner's handbook. Guilford Press.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*(1), 72–110. https://doi.org/10.2307/1170287

Vanderwood, Mike L., et al. "Predictive accuracy of nonsense word fluency for English language learners." *School Psychology Review*, vol. 37, no. 1, Mar. 2008, pp. 5+.

Rayner K, Foorman BR, Perfetti CA, Pesetsky D, Seidenberg MS. How psychological science informs the teaching of reading. Psychol Sci. 2001 Nov;2(2 Suppl):31-74. doi: 10.1111/1529-1006.00004.. PMID: 11878018.

Taylor, R. L. (2009). Assessment of exceptional students: Educational and psychological procedures (8th ed.) Upper Saddle River, NJ: Pearson Education.

Torgesen, J. K. (2002). The Prevention of Reading Difficulties. Journal of School Psychology, 40, 7-26. http://dx.doi.org/10.1016/S0022-4405(01)00092-

Test Administration

Overview

The Literably Screener is a computer-administered assessment of reading skills designed to help educators screen and monitor for reading difficulties in grades K-5.

Effective universal screening requires that the instrument is administered and scored with fidelity. Teacher-administered instruments can be time-consuming to administer and score, and research suggests that some educators struggle with consistently and accurately administering, scoring, and entering data (Taylor, 2009). As a computer-administered assessment that standardizes the assessment experience for students and scores student responses automatically, Literably saves educators time while maintaining fidelity in administration and scoring.

This manual describes protocols for administering the Literably Screener to students.

Administration Timeline and Mode

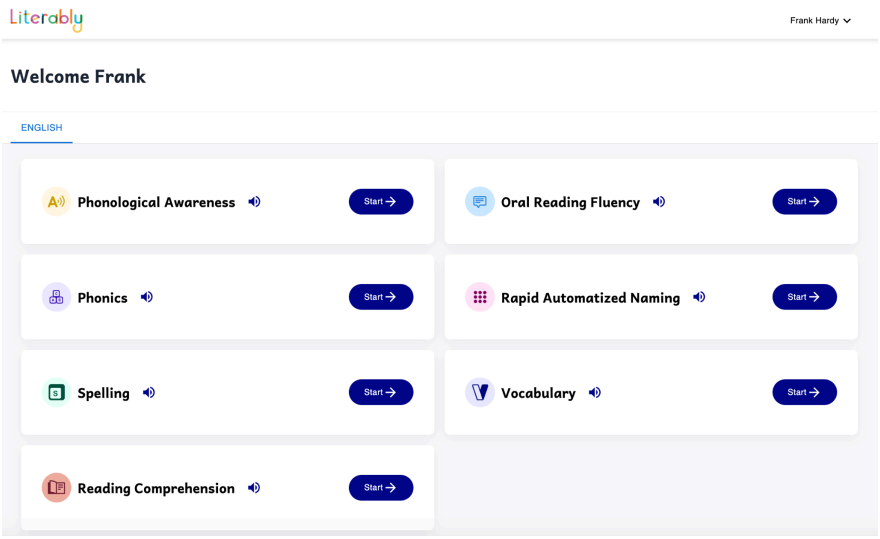
The Literably Screener is recommended for universal screening purposes during the following 3 windows:

- Fall: July 16 - December 31
- Winter: January 1 - March 31
- Spring: April 1 - July 15

The Literably Screener is administered via an online platform and can be completed on an internet-connected device that has speakers and a microphone, such as a laptop, desktop, Chromebook, or tablet. Full technical specifications are available in the Literably Help Center.

Prior to testing, students will have been rostered in the Literably database and assigned login credentials. Test administrators assign the appropriate screener to students by grade level and time of year. Students take the screener by logging into their Literably account using a username and password or via a single sign-on platform such as Clever or Classlink. The subtest options will appear on a choice menu, as shown in *Figure 16* below. Once a student completes a subtest, the option will disappear from the menu.

Figure 16. *Student Subtest Choice Menu*



Because the Literably Screener is a computer-administered assessment, all directions are delivered digitally through the Literably assessment interface. Students complete the assessment by providing oral responses, clicking a mouse or tapping a touchscreen, or typing letters on the keyboard. All student responses are collected by the platform and scored directly by Literably.

Technology Requirements

On laptops and desktops, Literably is compatible with most modern browsers and can be accessed via literably.com/login. It is recommended that devices use the most up-to-date versions of accepted browsers:

- Google Chrome 60 and newer (Flash not needed) - This is the recommended browser for Literably.
- Firefox 40 and newer (Flash not needed)
- Microsoft Edge Chromium (Flash not needed)

On iPads, students assess through the Literably app, which can be downloaded for free from the Apple App Store. App versions 1.3.2 or higher are supported, and iPadOS 13.0 or higher is required.

A reliable internet connection and the safelisting of required URLs within the network are required in order for Literably assessments to transmit successfully. External headsets with microphones are recommended for enhancing audio quality, but are not required. Literably needs to be granted permission to use the microphone at the start of the assessment session, and the student will be shown a pop-up alert to grant permission during their first login.

Before launching Literably across a school or district, it is recommended for school/district leaders and IT to review and complete the checklist below:

#	Action Item	Resources and Background Info.
1	"Safelist" Literably.	Literably's Safelist

2	Ensure students have access to the internet and have compatible devices to assess on.	Literably works on iPads (iPad OS 13.0 or above), laptops, Chromebooks, and desktops
3a	If students are using iPads, ensure that their iPads are compatible with Literably and that the latest Literably app has been pushed out to them (or ensure students have downloaded the Literably app from the App Store themselves).	Information about iPad compatibility and the Literably iPad app
3b	If students are using computers, ensure they are using a compatible internet browser (Google Chrome, Firefox, or Edge Chromium).	Information about compatible internet browsers
4	(Optional) If students are using headsets/external microphones, please make sure those pieces of equipment are available and working correctly.	
5	Upload/share the school's or district's rosters on Literably (whether IT is responsible for rostering will vary by school/district).	Information about Literably's rostering options
6	Before instructing teachers/students to start assessing, please test out logging in as a student, teacher, and admin to make sure everything is working correctly.	
7	Communicate to the necessary staff and students about how they will log into Literably (e.g., let users know what passwords they will be using to log into Literably). This action step can be completed by staff outside of IT.	Literably's customizable communication tools (emails, videos, etc.)

Administration Protocol

Table 54 summarizes the average administration time for each Literably subtest, along with the scoring entity, method, and time. Note: Literably's data suggests that average administration time does not significantly vary by subgroup.

Table 54. *Literably subtest administration and scoring time*

Literably Subtest	Administration Time (Minutes)	Scoring Entity, Method, and Time
Phonological Awareness	10 minutes	Scoring is done by trained Literably graders within 24 hours.
Phonics	5 minutes	Scoring is done by trained Literably graders within 24 hours.

Literably Subtest	Administration Time (Minutes)	Scoring Entity, Method, and Time
Rapid Automatized Naming	5 minutes	Scoring is done by trained Literably graders within 24 hours.
Vocabulary	5 minutes	Scoring is done immediately and automatically by Literably.
Oral Reading Fluency	10 minutes	Scoring is done by trained Literably graders with support from automatic speech recognition within 24 hours.
Spelling	5 minutes	Scoring is done immediately and automatically by Literably.
Reading Comprehension	10 minutes	Scoring is done immediately and automatically by Literably.

Table 55 below summarizes the average total administration time at each grade.

Table 55. *Literably Screener administration time by grade*

Grade	Average Administration Time (Minutes)
Kindergarten	25 minutes
1	40 minutes
2	35 minutes
3	35 minutes
4	30 minutes
5	30 minutes

The sections below describe the specific administration guidelines that pertain to each screener subtest, including number of items, assessment time, and assessment procedures. For each subtest, Literably suggests a recommended assessment grouping, which should be appropriate for the majority of students, as well as an optional grouping that may be more appropriate for students who need additional support. These may include very young students, ELs who typically receive accommodations in the classroom, or students who require small-group or one-on-one assessment as prescribed by their IEPs.

RAN

Grade Levels Administered	K, 1
Number of Times Per Year	Up to 3 (beginning, middle, end)

Number of Items Per Form	2 (Numbers, Letters)
Average Assessment Time	5 minutes
Assessment Grouping	Recommended: Small Group (6 maximum) Optional: One-on-One

To take the RAN subtest, students log into the Literably platform and follow oral prompts delivered by an animated character and visual prompts on the screen. The screen will display 2 different arrays of 50 stimuli: numbers and letters. Students are given the names of the stimuli before each task, along with a short practice session. When prompted with each array, students will name the stimuli in the array as quickly as they can, before moving to the next array. Students need to press a green button to submit each oral response.

Students are recommended to assess in a small group of not more than 6 students, spaced at least two feet apart. The test administrator should ensure that students are able to successfully login and navigate the assessment, and that the microphones are functioning on student devices. For students who need additional support, it is recommended that the test administrator sit one-on-one with the student to assist with navigating the platform.

Headsets with built-in microphones are recommended to reduce background noise and improve the audio quality of student responses, but they are not required. Students can pause and resume the assessment in between tasks.

Phonological Awareness

Grade Levels Administered	K, 1
Number of Times Per Year	Up to 3 (beginning, middle, end)
Number of Items Per Form	20
Average Assessment Time	10 minutes
Assessment Grouping	Recommended: Small Group (6 maximum) Optional: One-on-One

To take the Phonological Awareness subtest, students log into the Literably platform and follow oral prompts delivered by an animated character. Students are given a sample question and practice item at the beginning of each task. For example, before completing phoneme segmentation items, students will hear an example and complete a practice phoneme segmentation question.

When prompted, students provide a verbal response. Students do not see questions on the screen, and they need to press a green button to submit each oral response. Students have 10 seconds to answer each prompt before being moved to the next question.

Students are recommended to assess in a small group of not more than 6 students, spaced at least two feet apart. The test administrator should ensure that students are able to successfully login and navigate the assessment, and that the microphones are functioning on student

devices. For students who need additional support, it is recommended that the test administrator sit one-on-one with the student to assist with navigating the platform.

Headsets with built-in microphones are recommended to reduce background noise and improve the audio quality of student responses, but they are not required. Students can pause and resume the assessment.

Phonics

Grade Levels Administered	K, 1, 2, 3
Number of Times Per Year	Up to 3 (beginning, middle, end)
Number of Items Per Form	25 in Grade 1, 20 in all other grades
Average Assessment Time	5 minutes
Assessment Grouping	Recommended: Small Group (6 maximum) Optional: One-on-One

To take the Phonics subtest, students log into the Literably platform and follow oral prompts delivered by an animated character and visual prompts on the screen. The screen will display letters or words individually. When prompted, students will provide a verbal response (e.g., the name of the letter or their reading of the word). Students need to press a green button to submit each oral response. Students have 10 seconds to answer each prompt before being moved to the next question.

Students are recommended to assess in a small group of not more than 6 students, spaced at least two feet apart. The test administrator should ensure that students are able to successfully login and navigate the assessment, and that the microphones are functioning on student devices. For students who need additional support, it is recommended that the test administrator sit one-on-one with the student to assist with navigating the platform.

Headsets with built-in microphones are recommended to reduce background noise and improve the audio quality of student responses, but they are not required. Students can pause and resume the assessment.

Spelling

Grade Levels Administered	1, 2, 3, 4, 5
Number of Times Per Year	Up to 3 (beginning, middle, end)
Number of Items Per Form	20
Average Assessment Time	5 minutes
Assessment Grouping	Recommended: Whole Group Optional: Small Group or One-on-One

To take the Spelling subtest, students log into the Literably platform and follow oral prompts delivered by an animated character. Students will be prompted to spell words using the

on-screen keyboard or the keyboard on their device. After spelling each word, students will click a green button to submit their response and move to the next item. There is no time limit. Before starting the assessment, students will see an example question and how to type the response.

Because Spelling does not require oral responses from students, it can be administered whole-class, with each student completing the assessment independently. It is recommended that students use headphones to minimize distractions from other students' devices. For students who need additional support navigating the platform, the assessment can be administered in a small group or one-on-one. Students can pause and resume the assessment.

Oral Reading Fluency

Grade Levels Administered	1, 2, 3, 4, 5
Number of Times Per Year	Up to 3 (beginning, middle, end)
Number of Items Per Form	One passage, 5 multiple choice comprehension questions
Average Assessment Time	10 minutes
Assessment Grouping	Recommended: Small Group Optional: One-on-One

To take the Oral Reading Fluency subtest, students log into the Literably platform. Students will be presented with a reading passage on multiple pages, accompanied by pictures. Students will read the passage out loud and click the "Next Page" button to see each page of the text. After reading the entire passage, the student will stop recording, and the audio file will be sent to Literably for scoring. The student will then answer five multiple choice comprehension questions. Students can reference the text while answering these questions.

Students are recommended to assess in a small group of not more than 6 students, spaced at least two feet apart. The test administrator should ensure that students are able to successfully login and navigate the assessment, and that the microphones are functioning on student devices. For students who need additional support, it is recommended that the test administrator sit one-on-one with the student to assist with navigating the platform.

Headsets with built-in microphones are recommended to reduce background noise and improve the audio quality of student responses, but they are not required. The Oral Reading Fluency assessment cannot be paused and needs to be completed in one continuous session.

Vocabulary

Grade Levels Administered	K, 1, 2, 3, 4, 5
Number of Times Per Year	Up to 3 (beginning, middle, end)
Number of Items Per Form	20
Average Assessment Time	5 minutes
Assessment Grouping	Recommended: Whole Group Optional: Small Group or One-on-One

To take the Vocabulary subtest, students log into the Literably platform and follow oral prompts delivered by an animated character and visual prompts on the screen. For each item, students will be prompted to select the answer choice out of 4 possible answer choices. After selecting a response, students will click a green button to submit the response and move to the next item. Before starting each task (e.g., picture matching, synonym, context), students will see an example question and how to select the response. For any words displayed on the screen, students have the option of hearing them read aloud by pressing a speaker icon.

Because Vocabulary does not require oral responses from students, it can be administered whole-class, with each student completing the assessment independently. It is recommended that students use headphones to minimize distractions from other students' devices. For students who need additional support navigating the platform, the assessment can be administered in a small group or one-on-one. Students can pause and resume the assessment. The assessment is not timed.

Reading Comprehension

Grade Level Administered	2, 3, 4, 5
Number of Times Per Year	Up to 3 (beginning, middle, end)
Number of Items Per Form	2 passages with 6 multiple-choice comprehension questions per passage
Average Assessment Time	10 minutes
Assessment Grouping	Recommended: Whole Group Optional: Small Group

To take the Reading Comprehension subtest, students log into the Literably platform and follow directions on the screen. Students have the option to have the directions read aloud by the platform.

Students read 2 grade-level passages silently. After each passage, they answer 6 multiple-choice questions. Students can return to the text while answering the questions. After moving on to the next text, students cannot return to the questions for a previous text.

Because Reading Comprehension does not require oral responses from students, it can be administered in a whole group, with each student completing the assessment independently. For students who need additional support navigating the platform, the assessment can be administered in a small group. Students can pause and resume the assessment, and the assessment is not timed.

Testing Irregularities

With the exception of Oral Reading Fluency, all Literably Screener subtests can be paused and resumed. If an interruption occurs that requires the student to stop and resume an assessment, the student can either pause or log out of the session, or be automatically logged out if the device is shut down. When the student logs back in, the assessment can be resumed at the same location.

For Literably Screener subtests that require oral responses from the student (Phonological Awareness, Phonics, RAN, and Oral Reading Fluency), occasionally the audio recordings cannot be graded. The most common reasons are excessive background noise, a defective device microphone, or the student speaking too quietly to be understood. If a teacher receives an unscorable result for one of the subtests, it is recommended that the student retake the subtest. In order to generate a composite score, the student needs to have a score from each applicable subtest at the student's grade level. If a student accidentally takes a subtest twice within the same testing window and produces two scores, Literably will use the first subtest score to calculate the composite score.

In the event of other testing irregularities that may impact test validity, educators can consult with district personnel or contact support@literably.com for more specific guidance.

Test Administrators

The Literably assessment platform contains embedded directions and requires students to proceed through the assessment in an established sequence. This ensures that administration is standardized and makes it relatively easy for test administrators to oversee and monitor the completion of the Literably Screener.

The Literably Screener can be administered by classroom teachers or other personnel at school sites who are responsible for working with students, such as literacy coaches or reading specialists. It is recommended for test administrators to receive a one-hour synchronous webinar training or watch an asynchronous recorded training video (described under Training and Resources below) to become familiar with the Literably assessment platform. Beyond this, assessors do not require specific qualifications or technological expertise.

During the administration of the assessment, test administrators are expected to complete the following responsibilities:

- Before assessing, use the Demo Student feature to explain the assessment platform and model how to navigate through the subtests (as needed)
- Ensure that students have the appropriate devices and are able to log into their accounts
- Provide a reasonably calm testing environment
- Supervise administration to ensure on-task behavior and assist any students who need support navigating the assessment
- In some cases, sit with a student one-on-one to help the student navigate the assessment (as needed)
- Check that all assessments have been completed by students

Trainings and Resources

Literably provides comprehensive and hands-on support and professional development throughout its district and school partnerships.

When a district adopts Literably, the Customer Success Team meets with district leadership to carry out the onboarding and setup process, which includes rostering, technology requirements, the determination of training needs, and other steps to ensure smooth implementation.

Literably training options are flexible, targeted, and engaging. Trainings can be delivered in the following formats:

- Synchronous live webinar training (most common)
- Asynchronous pre-recorded training
- In-person live training (by request)

There are three standard training courses available:

- *Literably 101 - Intro to Literably (1 hour):* This training introduces users to the Literably platform and provides an overview of the Literably Screener. The training covers topics such as site navigation, how to administer the assessment to students, how to interpret results, and how to communicate to parents/guardians about screener results. The Literably 101 training is included in the implementation cost for all first-year adopters of Literably, and can be purchased as an add-on in later years. It is recommended that all staff members who will be administering the Literably Screener participate in this training in order to successfully assess students in Literably.
- *Literably 201 - Advanced Topics (1 hour):* This training is designed for staff members who have already used Literably with students and are looking for tools to dig more deeply into their data, in order to make instructional next steps informed by Literably results. This hands-on session leads teachers in examining their dashboards to analyze their student data, incorporating group discussion and independent work time.
- *Literably 301 - Literably for Administrators (1 hour):* This training is geared towards district and school administrators using Literably. It focuses on examining Literably data from a district- or school-level perspective and leveraging various Literably data reports to inform decision-making.

Literably can also design custom training sessions in collaboration with district partners to meet their specific needs.

In addition to the training options above, Literably offers the following forms of support at no additional cost to educators:

- *Help Center:* The Literably Help Center is a comprehensive knowledge base containing articles and how-to videos that help educators get the most out of Literably. It is directly accessible from the teacher dashboard.
- *Support Center:* The Literably Support Center is a help desk that provides email support to users within 24 hours on all topics related to using Literably. Teachers and administrators can submit questions directly from their dashboards.
- *Customer Success Manager:* Every Literably account has a dedicated Customer Success Manager who can assist via email, phone, or Zoom with every aspect of implementing Literably.

Accessibility Features

Literably offers accessibility solutions to respond to the unique needs of partner schools, including student impediments to testing (e.g., disability, language), and can support districts in administering assessments as modified. Some solutions are embedded in Literably's technology, and others are non-embedded (provided by schools non-digitally or via third-party software).

Literably's accessibility solutions fall into three categories:

- *Universal features* are provided to all students to reduce barriers to valid measurement.

- *Designated features* can be provided at the educator(s)' discretion.
- *Accommodations* can be provided to students with a documented need (e.g., on an Individualized Education Program (IEP) or 504 accommodation plan).

Universal and designated features are described during the Literably 101 training, and information regarding all accessibility features can be found in resources provided to districts during the onboarding process. When deciding whether to make a designated feature or accommodation available to a student, schools should follow their state accessibility guidelines.

Literably's universal features, designated features, and accommodations are described in *Tables 56-58*.

Table 56. *Universal features*

Feature	Description	Subtest(s)	Embedded?
Teacher Modeling	Teachers can model how to navigate the assessment platform using the Demo Student. It is recommended that all students receive exposure to Literably via this tool before their first time taking assessments.	All	Yes
Audio instructions	The student can click to hear any instruction read aloud.	All	Yes
Audio items	The student can click to hear a question or answer choice read aloud.	Vocabulary, Reading Comprehension	Yes
Volume control	The student can adjust the volume. This can be used to support students who are hard of hearing.	All	Yes
Pause	The student can take breaks during subtests by pausing and resuming the assessment. Breaks naturally occur between subtests.	All except for Oral Reading Fluency	No
Tracking aid	The student can use a tracking aid (e.g., a marker or ruler).	All	No
Noise buffer	The student can use a noise buffer (e.g., headphones) to minimize distraction.	All	No
Standard assistive technology	The student can use assistive technology they typically use to access content (e.g., hearing aids, glasses).	All	No
Note-taking	The student can make notes on scratch paper, an erasable whiteboard, or an approved assistive technology device. All	Reading Comprehension	No

	notes must be securely destroyed.		
Grouping	To limit distractions for students and provide individualized support, students can be assessed one-on-one or in small groups (6 or fewer).	All	No

Table 57. *Designated features (Non-Embedded)*

Feature	Description	Subtest(s)
Separate setting	The school can designate a separate test location for students who require minimal distractions.	All
Color contrast	The student can use color contrast technology (e.g., Chromebook, Windows, Mac, iPad).	All
Magnification device	The student can use assistive technology to adjust the size of specific areas of the screen (e.g., ZoomText). On iPads, students can pinch to zoom in on the screen. During one-on-one administration, the assessment can be projected onto a larger screen.	All
Human navigator	A human navigator can help the student navigate the test, including reading instructions aloud.	All
Human item reader	A human reader can read the item content aloud.	Reading Comprehension
Human translator	A test administrator fluent in the student's native language can translate audio instructions into the student's native language	All

Table 58. *Accommodations (Non-Embedded)*

Accommodation	Description	Subtest(s)
Specialized assistive technology	The student can use specialized assistive technology (e.g. customized keyboards; customized pointing devices) to complete the test.	All
Human signer	A test administrator fluent in sign language can sign audio instructions.	All
Refreshable braille	Students who are visually impaired can use a refreshable braille device.	Phonics, Oral Reading Fluency, Vocabulary, Reading Comprehension

Human scribe	The student can dictate their responses to a human scribe who records the student's response. This can be used to support students with limited fine motor skills.	Vocabulary, Reading Comprehension
Screen reader	The student can use screen reader technology to identify what is displayed on the screen.	Vocabulary, Reading Comprehension
Visual Cue	Administrator provides a visual cue to the student (e.g., visual phonics, lip-speech reading, cued speech, etc.) to support students who are hard of hearing.	Phonics, Phonological Awareness, Spelling

In addition to the universal features, designated features, and accommodations described above, Literably can further modify the Literably Screener based on the unique needs of partner schools on a case-by-case basis when requested modifications would not render the assessment invalid.

Table 59 below outlines special considerations that can help to guide the use of these accessibility features for different student populations.

***Table 59**. Special Considerations*

Student Population	**Considerations**
English Language Learners	English language learners do not require accommodations to take the Literably Screener, but for specific students, it may be appropriate to provide access to the following designated features: separate setting, human navigator, human translator.
Students who are deaf or hard of hearing	The following accessibility features may be appropriate for students, in accordance with each student's IEP: volume control, noise buffer, standard assistive technology (e.g., hearing aids), human signer, visual cue. Before testing, students' auditory listening devices should be in good working condition, and students should be seated close to the test administrator. If appropriate, students may be assessed by a speech language pathologist or teacher who regularly works with the student.
Students with visual impairments	The following accessibility features may be appropriate for students, in accordance with each student's IEP: standard assistive technology (e.g., glasses), color contrast, magnification device, refreshable braille, screen reader.
Students with severe disabilities	The following accessibility features may be appropriate for students, in accordance with each student's IEP: separate setting, human navigator, human item reader, specialized assistive technology, human scribe.

Description of Test Security Measures

Confidentiality of Student Data

Literably's Privacy Policy, which can be accessed on the Literably website and via this link, describes safeguards to ensure the confidentiality of student data.

These safeguards are summarized below:

When Literably is used by a school, we may collect or have access to student data, including personal information from a student's educational records as defined by FERPA. Literably considers student data to be highly confidential and does not use such data for any purpose other than providing our service to the school and as otherwise provided in our agreements with the school. Our collection, use, and disclosure of student data is governed by our Terms of Service and/or any other agreement with the school, by the provisions of the Family Educational Rights and Privacy Act ("FERPA"), the Children's Online Privacy Protection Act ("COPPA") and applicable state laws relating to the collection and use of personal information of students.

In addition to Literably's Privacy Policy, Literably enters into Data Privacy Agreements, or DPAs, with districts based on local requirements, to maintain compliance with state and national laws governing student data privacy.

Assessment Integrity

Table 60 describes the test security measures in place to safeguard the integrity of Literably assessment items.

Table 60. *Test Security Measures*

	Description
Role-Based Access	● Educators and students are rostered through secure system applications. ● Literably users are only granted the level of account access necessary, in relation to their role within their district or school. ● Only administrator accounts with rostering privileges have the ability to create or modify student profiles. ● Parents can only view their child's assessment data using a unique "secret access link" to that student, which can be provided by the school.
Item Security	● Literably assessment items are only accessible to authorized users. ● Every student must log into his/her own Literably account and can only complete assessments assigned by the teacher. ● Students can only have one assessment session open at one time. ● Student access can be limited to school-hours on weekdays, to allow assessment access only under direct supervision of school personnel.

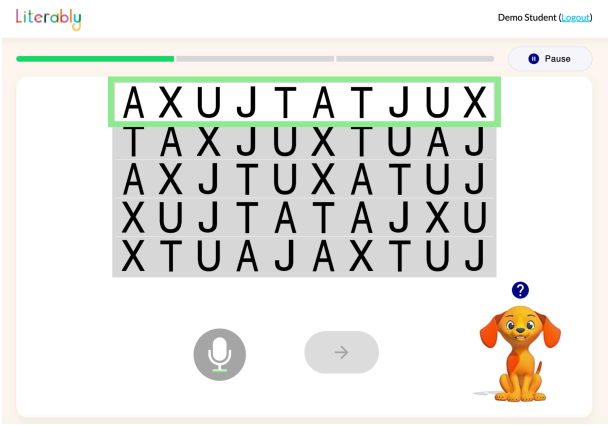
	● Literably will not assign the same oral reading fluency passage to a student after the student has assessed on that passage, and the assessment has been scored. ● Literably test items are stored on secure servers and are not stored or cached locally.
Data Transmission and Storage	● Literably assessment data are stored in highly-secure Amazon data centers throughout the continental U.S. ● Transmission of student response data and personally identifiable information is encrypted through the HTTPS encryption protocol. ● Student response data and PII are stored in an encrypted database at rest, and raw passwords are not stored.

Sample Test Forms and Score Reports

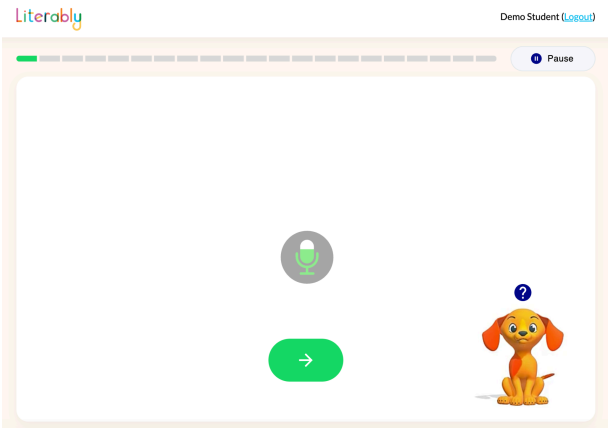
Sample Test Items

The tables below contain sample test items from the Literably Screener subtests, including the visual prompt presented to the student and a description of the task.

RAN Subtest

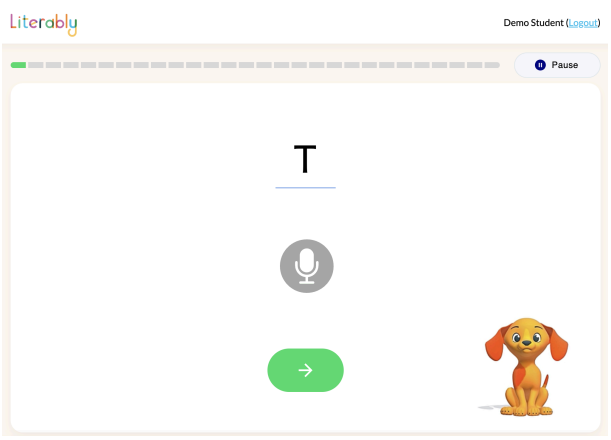
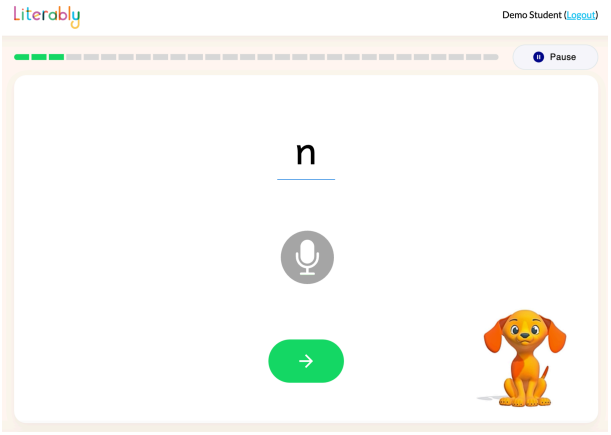
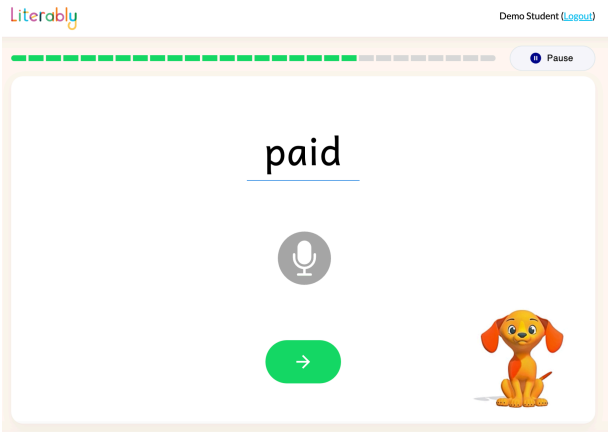
Skill / Grade	Screen Display	Description
Rapid Automatized Naming - 1st Grade		For each RAN task, students will name 50 items in a 10x5 array and press the green arrow to move to the next task. The image to the left shows the RAN Letters task. Literably RAN also includes Numbers.

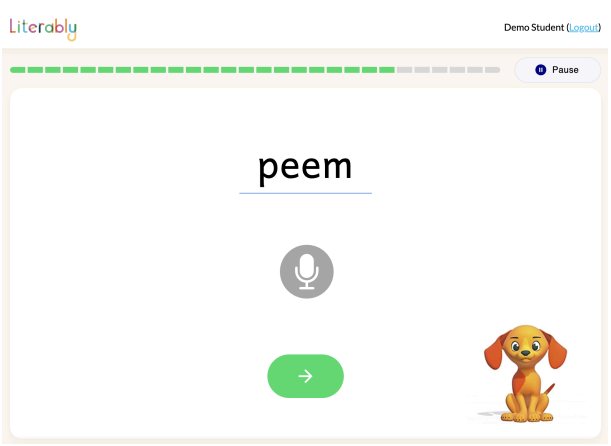
Phonological Awareness Subtest

Skill / Grade	Screen Display	Description
Phoneme Segmentation - 1st Grade		Rudy (animated dog) delivers each prompt to the student. For example: “Say ‘milk.’ Say the sounds you hear in ‘milk.’”

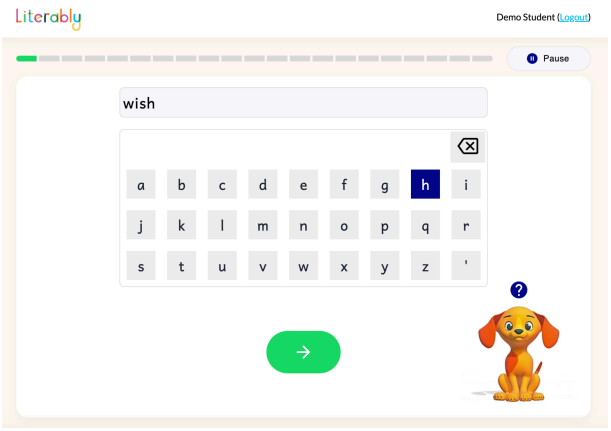
Phonics Subtest

Skill / Grade	Screen Display	Description

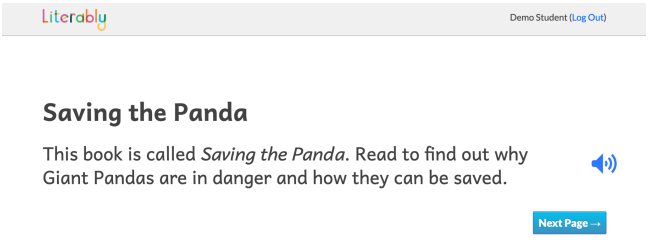
Letter Name - 1st Grade		Rudy asks the student: “What is the name of this letter?”
Letter Sound - 1st Grade		Rudy asks: “What sound does this letter make?”
Word Reading - 1st Grade		Rudy prompts the student: “Read the word on the screen out loud.”

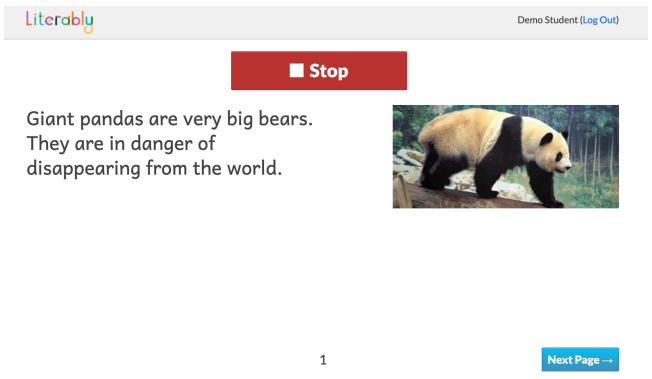
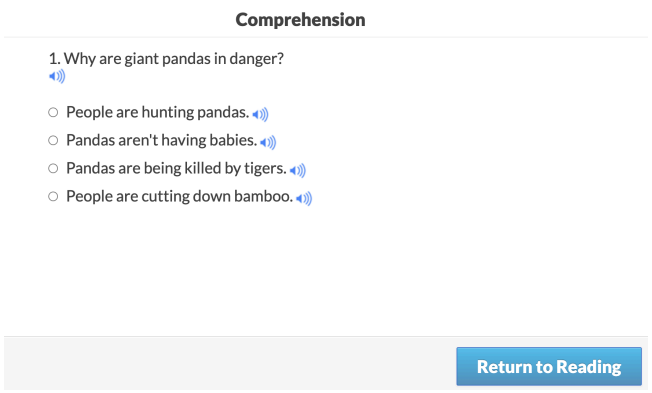
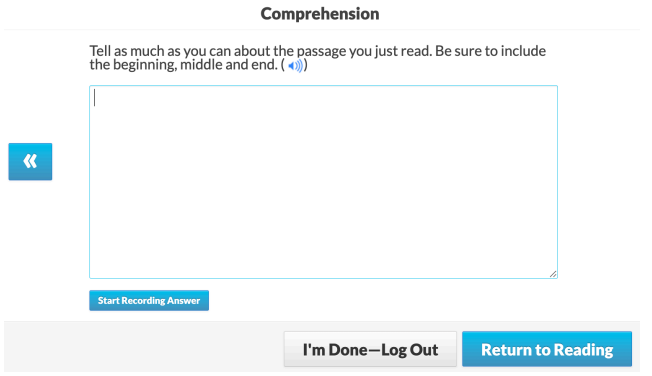
Nonsense Word Reading - 1st Grade		Rudy prompts the student: “Try your best to read this make-believe word.”

Spelling Subtest

Skill / Grade	Screen Display	Description
Spelling - 1st Grade		Rudy provides the prompt: “Wish. I made a wish on my birthday. Spell wish.” The student can use the keyboard on the screen or on their device to spell the word.

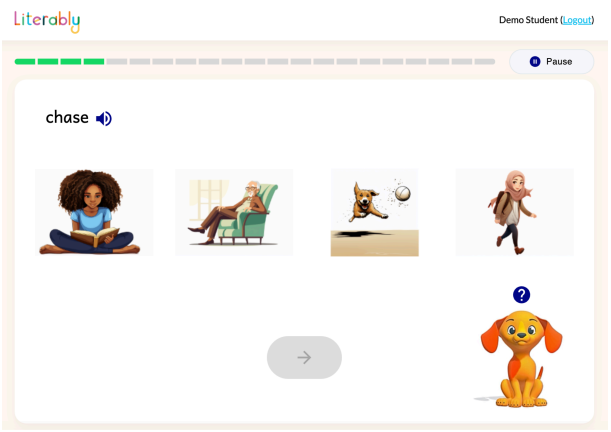
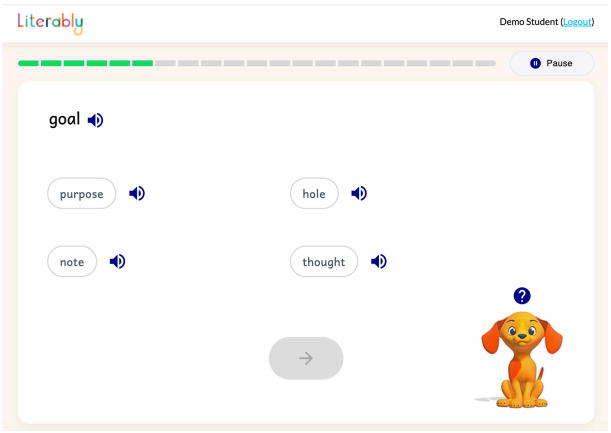
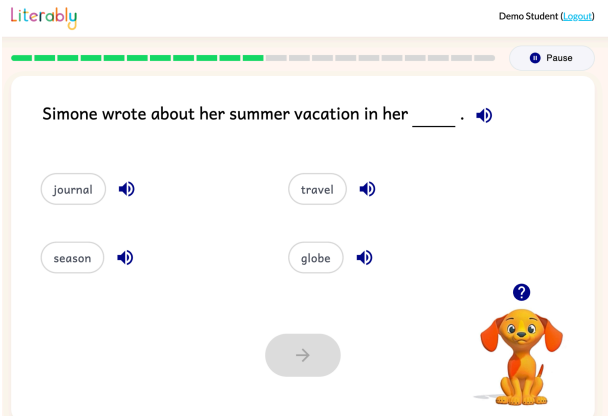
Oral Reading Fluency Subtest - 1st Grade

Test Segment	Screen Display	Description
Book Intro		Students are first presented with a brief description of the passage they will be reading for the assessment.

Oral Reading Recording		Students press the blue “Start” button to begin recording and press “Next Page” to move through the pages. At grades 1-3, passages have accompanying pictures. At grades 4+, texts do not have pictures.
Multiple-Choice Comprehension Questions		Each passage has five multiple-choice comprehension questions.
Retell		Students can type or speak an optional retell response.

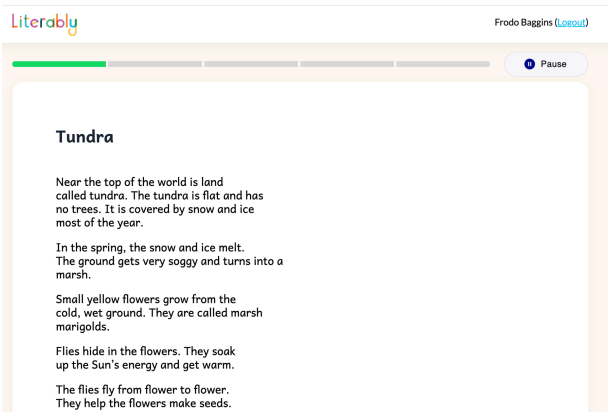
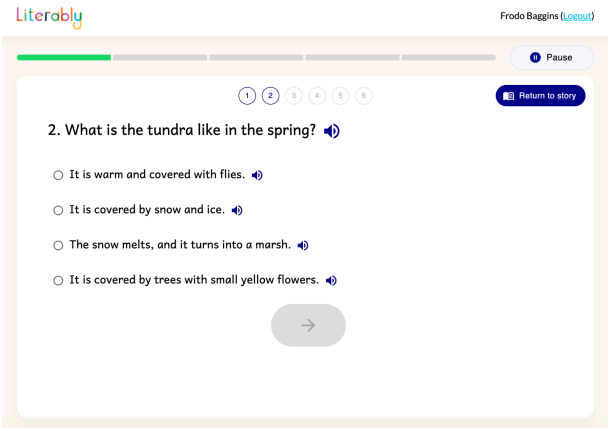
Vocabulary Subtest

Skill / Grade	Screen Display	Description

Picture Matching - 1st Grade		Rudy asks, “Which picture best matches the word.” Students select the best answer out of 4 choices.
Synonym - 1st Grade		Rudy asks, “Which word has the same or similar meaning?” Students select the best answer out of 4 choices.
Context - 1st Grade		Rudy asks, “Which word best completes the sentence.” Students select the best answer out of 4 choices.

Reading Comprehension Subtest

Test Segment	Screen Display	Description

Reading Passage - 2nd Grade		Students read each passage silently.
Multiple-Choice Comprehension Questions		Students answer 6 questions for each text. There are 2 texts total.

Sample Score Reports

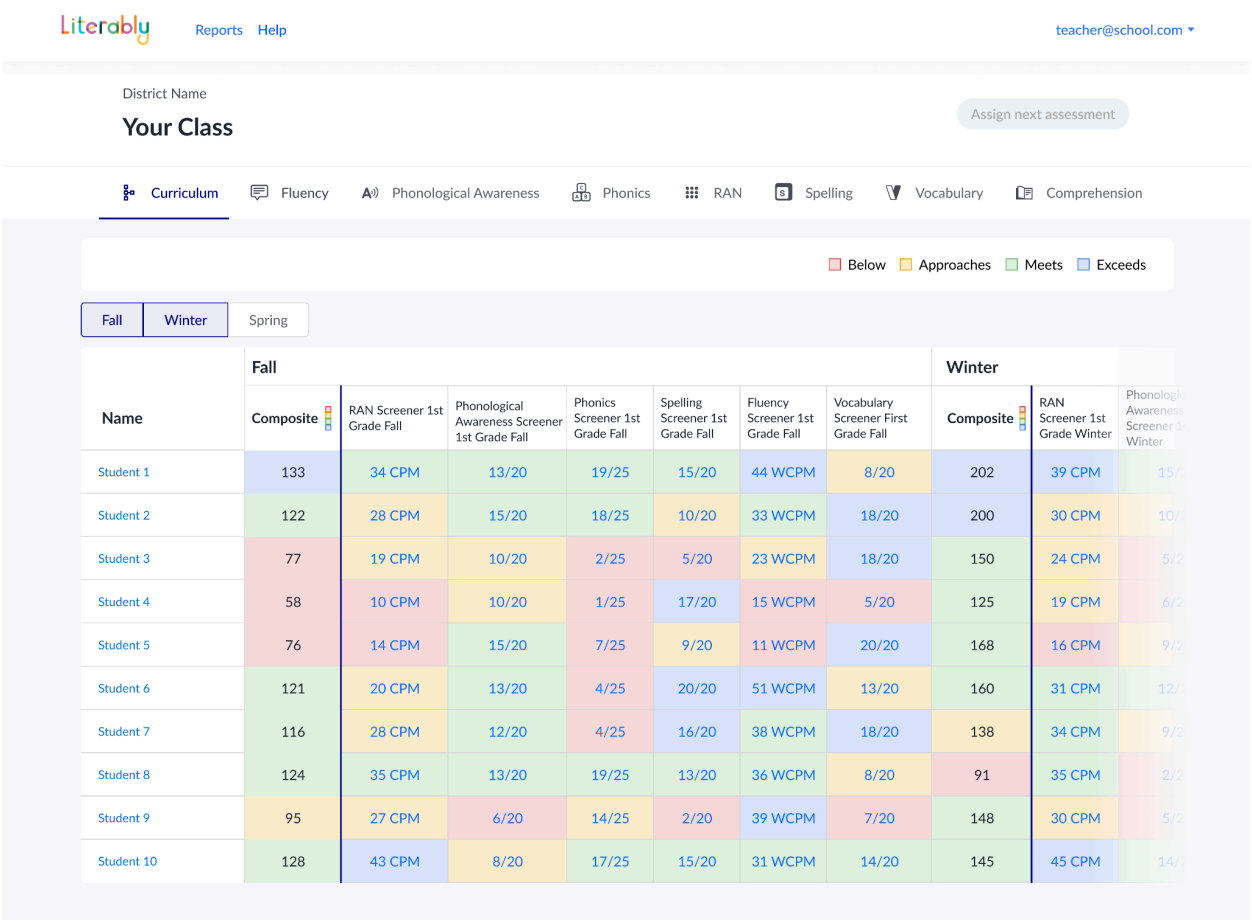
Reports

Using the Literably interface, educators can view Literably Screener data at the class, student, and assessment levels. Data can be exported from the Literably platform for further analysis, and districts also have the option to have data exports sent nightly via a secure SFTP server, to be ingested by their data management system of choice.

Classroom View

The classroom view allows teachers to view results from the Literably Screener for the entire class. This view displays composite scores and subtest scores for each student during each administration of the screener. The Composite Score is color-coded to indicate performance categories, and the columns can be sorted to view student groupings by performance.

Figure 17. *Classroom View*

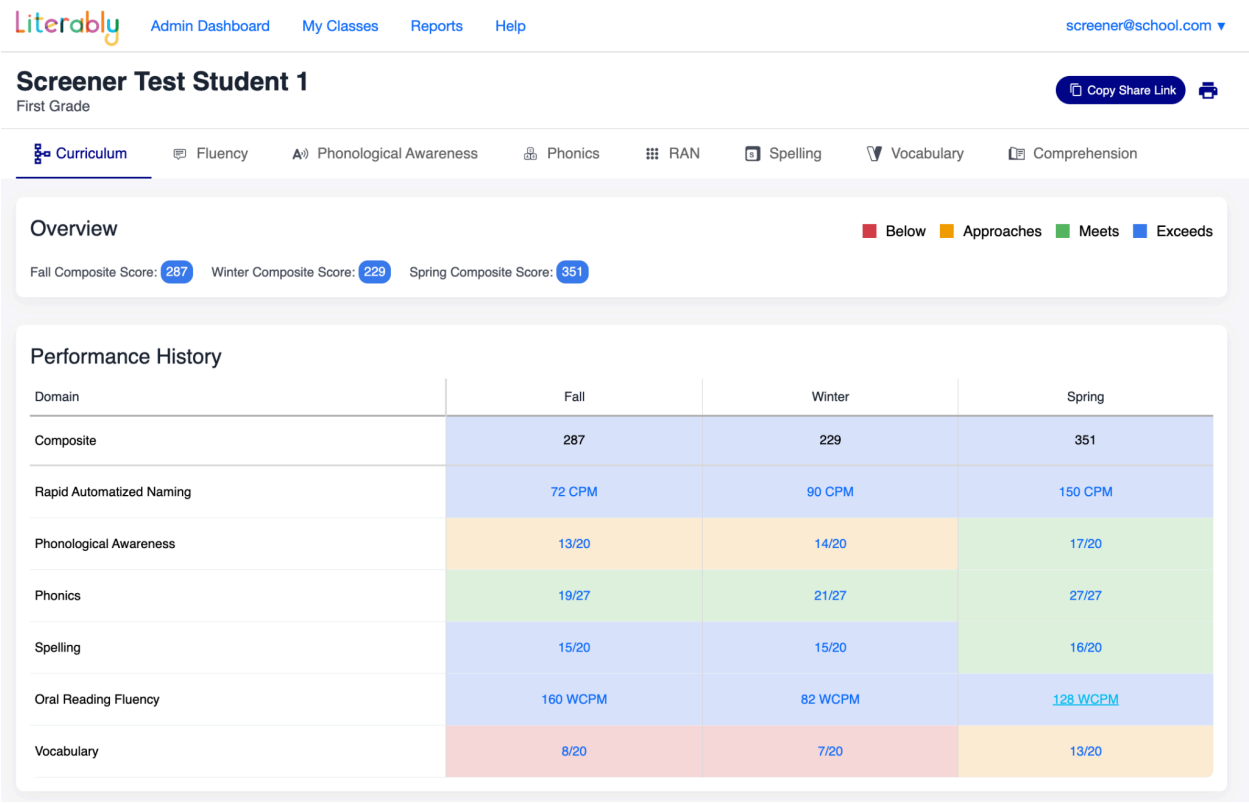


Name	Fall							Winter		
	Composite	RAN Screener 1st Grade Fall	Phonological Awareness Screener 1st Grade Fall	Phonics Screener 1st Grade Fall	Spelling Screener 1st Grade Fall	Fluency Screener 1st Grade Fall	Vocabulary Screener First Grade Fall	Composite	RAN Screener 1st Grade Winter	Phonological Awareness Screener 1 Winter
Student 1	133	34 CPM	13/20	19/25	15/20	44 WCPM	8/20	202	39 CPM	15/
Student 2	122	28 CPM	15/20	18/25	10/20	33 WCPM	18/20	200	30 CPM	10/
Student 3	77	19 CPM	10/20	2/25	5/20	23 WCPM	18/20	150	24 CPM	5/
Student 4	58	10 CPM	10/20	1/25	17/20	15 WCPM	5/20	125	19 CPM	6/
Student 5	76	14 CPM	15/20	7/25	9/20	11 WCPM	20/20	168	16 CPM	9/
Student 6	121	20 CPM	13/20	4/25	20/20	51 WCPM	13/20	160	31 CPM	12/
Student 7	116	28 CPM	12/20	4/25	16/20	38 WCPM	18/20	138	34 CPM	9/
Student 8	124	35 CPM	13/20	19/25	13/20	36 WCPM	8/20	91	35 CPM	2/
Student 9	95	27 CPM	6/20	14/25	2/20	39 WCPM	7/20	148	30 CPM	5/
Student 10	128	43 CPM	8/20	17/25	15/20	31 WCPM	14/20	145	45 CPM	14/

Student History View

The student history view displays composite and raw scores for each student on the screener, across seasons. This view can be printed or shared digitally using the Copy Share Link. A copy of this report can be sent home along with the Parent Letter to communicate screener results with families.

Figure 18. *Student History View*



Assessment View

For screening purposes, the Literably Screener Composite Score is sufficient for providing information about a student's level of reading risk and informing decisions related to instructional next steps. However, the Literably platform also allows educators to drill down into each subtest to view assessment results in more detail. These results pages report students' item responses and the corresponding skills addressed by each item. This granular information provides data that can be used beyond the screening process to analyze specific skill gaps within each domain.

Figure 19 shows a sample item performance section from the assessment view of the Phonics subtest. Student responses are displayed under the "Response" column, and teachers have the ability to listen to their students' oral responses.

Figure 19. *Assessment View - Phonics*

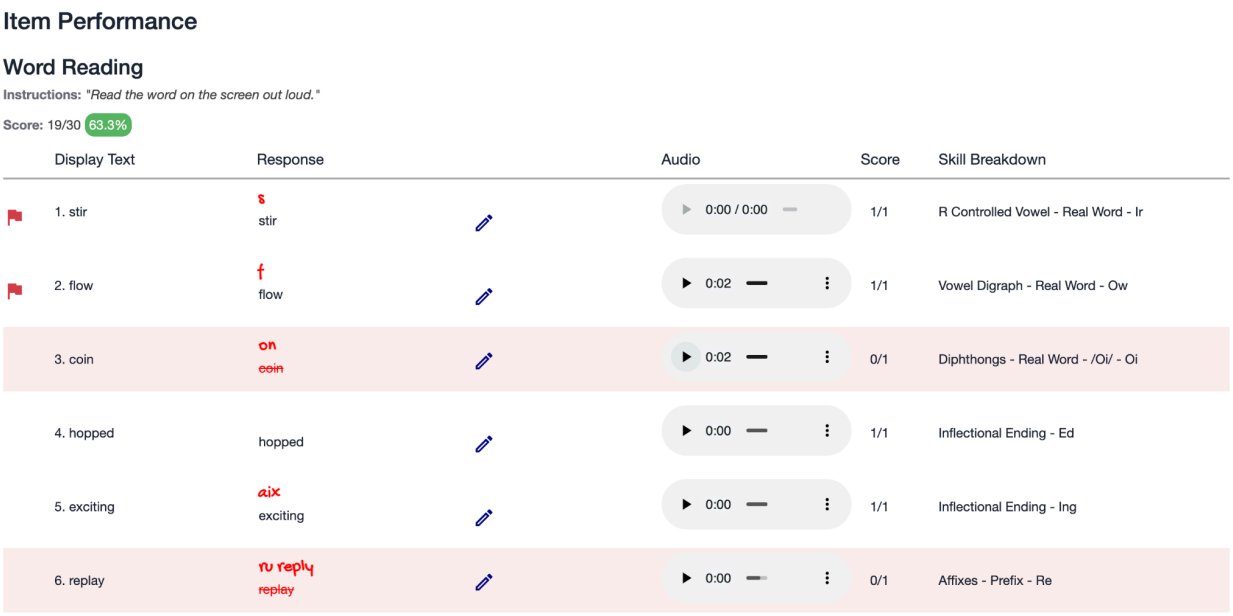


Figure 20 below shows a sample item performance section from the assessment view of a Vocabulary subtest, which displays the target words and answer choices from the picture matching section.

Figure 20. *Assessment View - Vocabulary*

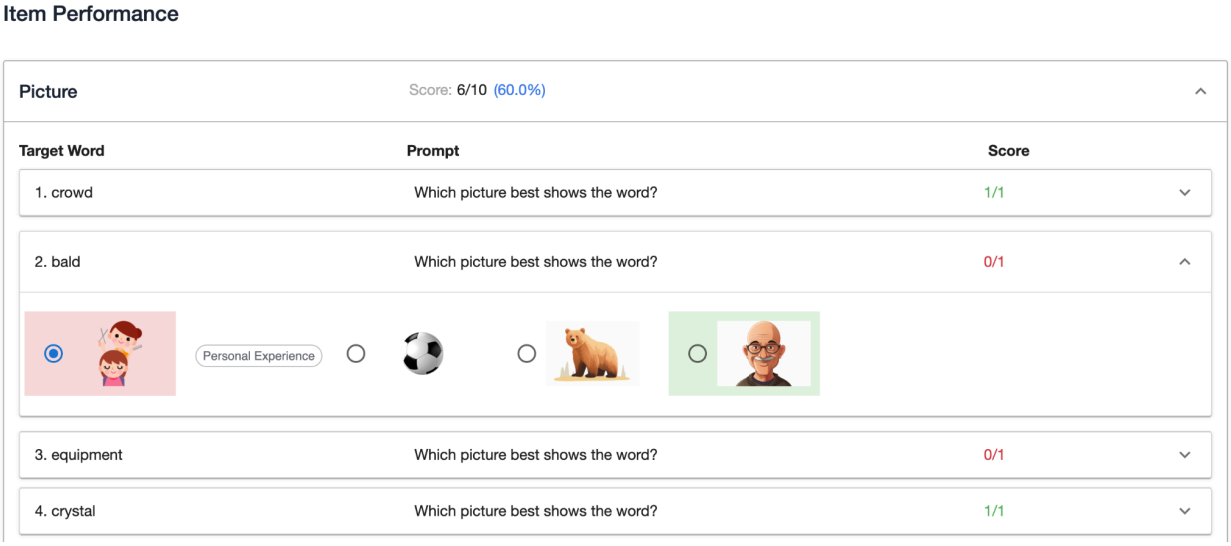
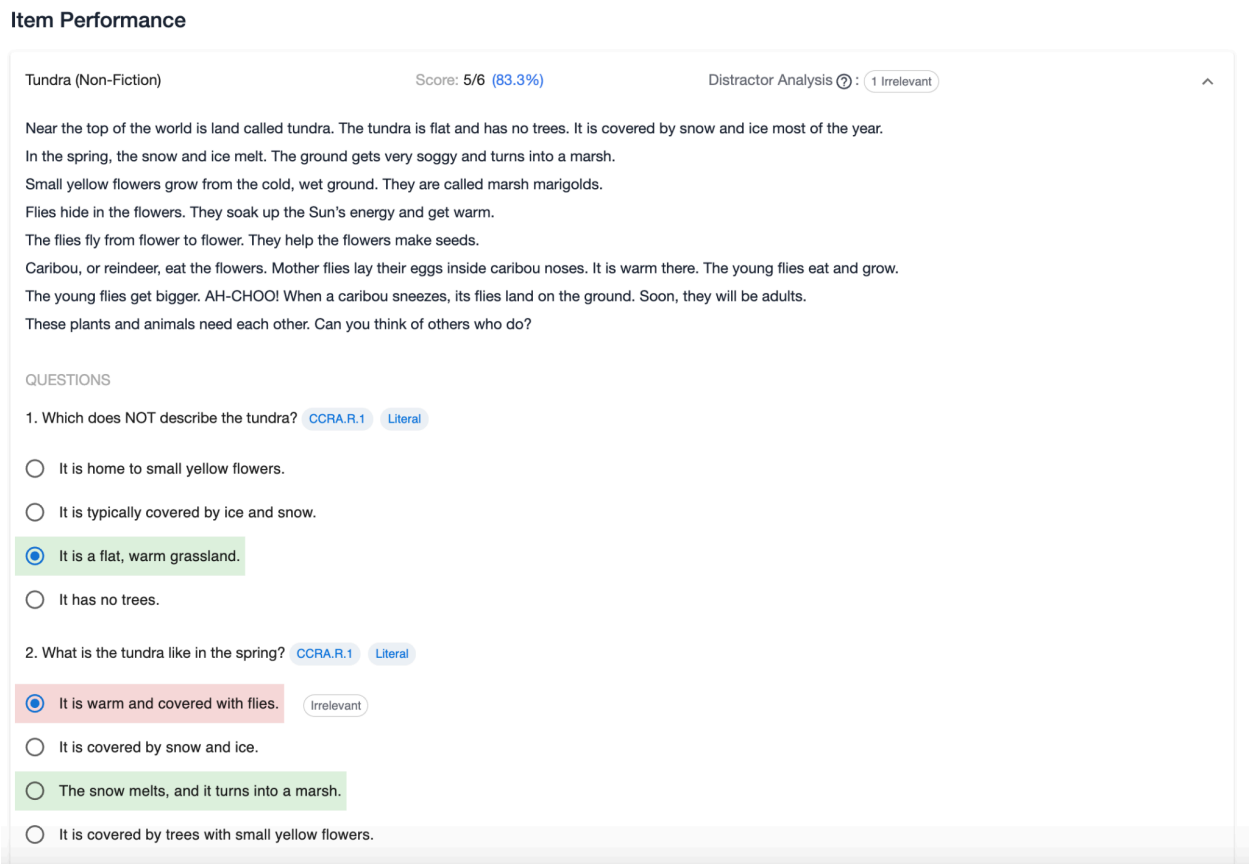


Figure 21 shows a sample item performance section from the assessment view of a Reading Comprehension subtest, which displays a reading passage, questions, and answer choices from the Reading Comprehension subtest.

Figure 21. *Assessment View - Reading Comprehension*



Data Exports

In addition to the data that is presented in the Literably user platform at the classroom, student, and assessment levels, described above, Literably also makes available a variety of data exports that can be downloaded by users with district or school administrator privileges. These reports are available at the district, school, and classroom levels to enable more detailed data analysis and decision-making, along with the ability to upload data to a data warehouse for reporting and archival purposes.

There are two main types of data exports:

1. Completion exports show how many students have completed screener assessments during each benchmark period, at the school and classroom levels. *Figure 22* below displays a sample of the completion export.

Figure 22. *Completion Report*

	A	B	C	D	E	F	G	H
1	Assessment Window	School	Grade	Assessed	Reassess Recommended	Total	Percent Assessed	Percent Needing to Reassess
2	1	Hogwarts School	2	86	4	97	88.70%	4.10%
3	2	Hogwarts School	1	95	2	100	95.00%	2.00%
4	2	Hogwarts School	2	113	4	118	95.80%	3.40%
5	3	Hogwarts School	1	90	3	95	94.70%	3.20%
6	3	Hogwarts School	2	115	1	118	97.50%	0.80%

2. Benchmark exports contain screener results for individual students, including subtest raw scores, composite scores, performance classifications, and student demographic information.

Parent Letter

In order to facilitate communication with families about Screener results, Literably provides a parent letter that describes the purpose, components, and results of the assessment. Literably recommends that teachers send this letter home after the completion of each screener administration, along with the student's assessment history (printed or linked) from the Literably platform (see Student History View above). If the student has been identified as needing additional intervention, it is recommended that school personnel meet with parents to discuss next steps.

The contents of the parent letter can be viewed below in English and Spanish.

Parent letter (English)

Student Name:
Date:

Dear Parent or Guardian,

Your child's school administers the Literably Screener. The purpose of the Literably Screener is to monitor students' development in reading, identify students who might need extra help, and to guide teachers in meeting students' learning needs.

The Literably Screener includes 7 short subtests that each measure important skills that contribute to proficient reading. Depending on your child's grade level, s/he may have completed some or all of these subtests. The subtests are listed and described below:

Rapid Automatized Naming (RAN) - Students name a series of familiar objects as quickly and accurately as they can, to show how well they can retrieve information.
Phonological Awareness - Students identify and manipulate sounds in words, to show that they can hear the sounds that make up words.
Phonics - Students name letters, give their sounds, and read grade-level words, to show their ability to blend letters into words.
Vocabulary - Students match words to pictures, identify synonyms, and select words to complete sentences, to show their knowledge of word meanings.
Spelling - Students spell grade-level words, to show their knowledge of phonics patterns.
Oral Reading Fluency - Students read a short passage aloud, to show they can read texts easily, quickly and accurately.
Reading Comprehension - Students read short passages silently and answer multiple-choice questions, to show they can understand what they read.

Each individual subtest cannot show the whole picture, but together they can tell a fuller story. The scores from the subtests are used to calculate a **Composite Score**, or the overall score your child received. The Composite Score gives schools one indication of whether the student is on track to meet grade-level expectations for reading.

Your child's teacher has attached a report of your child's scores on the most recent administration of the Literably Screener. The report shows the scores for each subtest and the

overall Composite Score. The Composite Score is color-coded, and the color indicates how your child performed in relation to grade-level expectations, based on the key below:

Color	Description
Blue	Your child's performance exceeds grade-level expectations.
Green	Your child's performance meets grade-level expectations.
Yellow	Your child's performance approaches grade-level expectations. Your child may need additional support and more frequent monitoring in order to meet grade-level reading goals.
Red	Your child's performance is below grade-level expectations. Your child will likely need additional instructional support and close monitoring in order to meet grade-level reading goals.

Depending on your child's results, your child's teacher may contact you to discuss next steps for supporting your child's learning. If you have any questions, please contact your child's teacher.

Thank you,

Parent letter (Spanish)

Nombre de el o la estudiante:
Fecha:

Querido padre, madre o tutor(a),

La escuela de su hijo(a) aplica el examen de evaluación de Literably. El propósito de esta evaluación es monitorear el desarrollo de la lectura de sus estudiantes, identificar a aquellos(as) que puedan necesitar ayuda adicional y guiar a las y los maestros en la satisfacción de las necesidades de aprendizaje de las y los estudiantes.

El examen de evaluación de Literably incluye 7 pruebas cortas que miden habilidades importantes que contribuyen a la lectura competente. Dependiendo del grado escolar de su hijo(a), es posible que haya completado todos o solo algunas de estas pruebas. Las pruebas se enumeran y describen a continuación:

Denominación Rápida Automatizada - Las y los estudiantes nombran una serie de objetos familiares tan rápida y precisamente como les sea posible para demostrar su capacidad para recuperar información.
Conciencia Fonológica - Las y los estudiantes identifican y manipulan sonidos en las palabras para mostrar que pueden escuchar los sonidos que componen las palabras.
Fonética - Las y los estudiantes nombran letras, enuncian sus sonidos y leen palabras apropiadas para su grado para demostrar la habilidad de transformar letras en palabras.
Vocabulario - Las y los estudiantes emparejan palabras con imágenes, identifican sinónimos y seleccionan palabras para completar las oraciones para mostrar su conocimiento del significado de las palabras.

Ortografía - Las y los estudiantes deletrean palabras adecuadas a su nivel de grado para mostrar su conocimiento de los patrones fonéticos.
Fluidez en la Lectura Oral - Las y los estudiantes leen un pasaje corto en voz alta para demostrar que pueden leer textos con facilidad, rapidez y precisión.
Comprensión Oral - Las y los estudiantes leen pasajes cortos en silencio y responden preguntas de opción múltiple para mostrar que pueden comprender lo que leen.

En lo individual, cada prueba nos muestra un lado de la historia, pero, al unirlas, podemos ver el panorama completo. Los resultados de cada prueba se usan para calcular una **Puntuación Compuesta** que su hijo(a) recibe. La Puntuación Compuesta da a las escuelas un indicador de si el o la estudiante está en camino de cumplir con las expectativas de lectura según su grado de lectura.

El maestro(a) de su hijo(a) ha adjuntado un reporte de los puntajes que su hijo(a) recibió en la última evaluación de Literably. El reporte muestra el puntaje de cada una de las pruebas y la Puntuación Compuesta general. La Puntuación Compuesta está codificada por colores y cada color indica cómo su hijo(a) se desempeñó en relación a las expectativas de su grado, basado en la siguiente clave:

Color	**Descripción**
Azul	El desempeño de su hijo(a) supera las expectativas de su nivel de grado.
Verde	El desempeño de su hijo(a) cumple con las expectativas de su nivel de grado.
Amarillo	El desempeño de su hijo(a) está cercano a las expectativas de su nivel de grado. Su hijo(a) puede necesitar apoyo adicional y un monitoreo más frecuente para cumplir con las expectativas de su grado escolar en cuanto a la lectura.
Rojo	El desempeño de su hijo(a) está debajo de las expectativas de su grado escolar. Su hijo(a) muy probablemente necesitará apoyo adicional y monitoreo cercano para estar a la par de las expectativas de su grado escolar en cuanto a la lectura.

Dependiendo de los resultados de su hijo(a), su maestro(a) puede contactarle para discutir los siguientes pasos para apoyar su aprendizaje. Si tiene cualquier pregunta, por favor contacte al maestro(a) de su hijo(a).

Gracias.

In addition, the Parent Letter is also available in the following languages:

- Vietnamese
- Chinese (Simplified)
- Arabic
- Russian
- Persian
- Filipino
- Punjabi
- Korean