English MAP Reading Fluency Technical Report

Based on assessments administered during the 2020–2022 school year

August 2024

nwea

Acknowledgments: It is with great appreciation that we recognize the many people at NWEA who contributed to this technical report. It was a collaborative effort involving people from numerous departments in the organization. We give special thanks to those who conducted the analyses and wrote and revised the document, including Carmen Hall, Wei He, Cindy Jiban, John Newburn, J. Patrick Meyer, Chuck Phillips, Kelly Rivard, Fang Peng, Mary Ann Simpson, Chris Wells, Kang Xue, and countless others whose expertise and knowledge about MAP Reading Fluency was crucial.

Table of Contents

List of Tables

List of Figures

Executive Summary

This technical report documents the processes and procedures employed by NWEA® to build and support the English MAP® Reading Fluency™ assessments administered during the 2020–2021 school year. It is written for measurement professionals and administrators to help evaluate the quality of MAP Reading Fluency. Principal information presented in each chapter is summarized below. This report is not intended to be an administration guide or a technical description of the hardware and software needed for use of the system. For additional information not covered in this technical report, please contact your local NWEA representative or consult the NWEA website at www.nwea.org.

Section 1: Introduction

This chapter presents an overview of the English MAP Reading Fluency assessment, including the rationale behind its design. MAP Reading Fluency is an adaptive online assessment that supports students on their path to reading comprehension by assessing and helping to improve both oral reading fluency and foundational reading skills. It is designed for students who do not yet read with solid fluency and understanding and adapts to accommodate pre-readers, early readers, and independent readers in pre-K to Grade 5 with the goal of helping all students be able to read with comprehension. NWEA began offering MAP Reading Fluency to the general public in Fall 2018, with progress monitoring being introduced in Fall 2019 and the dyslexia screener becoming available to all users in Fall 2021.

Section 2: Test Design

This chapter describes the MAP Reading Fluency test forms and provides the specifications for each measure included on the assessment. Teachers can choose from six test forms: Adaptive Oral Reading, Foundational Skills, Foundational Skills–Beginner, Passages Only, Dyslexia Screener, and Progress Monitoring. Adaptive Oral Reading is the default test form when no specific form is chosen. On the Adaptive Oral Reading form, students are routed to either Oral Reading Fluency or Foundational Skills based on their performance at the beginning of the assessment. The Oral Reading Fluency track presents students with passages, whereas the Foundational Skills track does not. The dyslexia screener collects data using foundational skills measures and a Rapid Automatized Naming (RAN) measure before using a predictive model to flag students. Progress monitoring provides a quick and reliable way to measure improvement in reading over time.

Section 3: Content Development

This chapter explains the process for developing items for the MAP Reading Fluency assessments. NWEA content specialists created item templates for each measure to ensure consistency in content scope, context, cognitive complexity, item format, graphics, and audio style. Stems were developed at the template level and were reviewed by experts in elementary grades for adherence to best practices for young students. Each item was written by NWEA content experts in elementary grades and received multiple reviews. Because stems were set at the template level, review at the item level focused on item assets (e.g., an audio and/or onscreen representation of a letter, sound, word, or sentence, possibly including a picture) and answer options (e.g., a letter, word, sentence, or picture, possibly with audio).

Section 4: Test Administration and Security

This chapter describes the test administration process, including setting up and managing students and test sessions. It also summarizes the test security procedures put in place by NWEA to ensure the integrity of the assessment and student information. MAP Reading Fluency is administered through the NWEA Comprehensive Assessment Platform. To take the assessment, each student needs a computing device and an over-ear headset with a boom microphone. All administration instructions are presented by audio within the test. The assessment experience uses two avatar contexts: Wiggles the worm for students in Pre-K to Grade 2 and Swift the yellow warbler for students in Grade 3 and above. Students typically take 20–30 minutes to complete the MAP Reading Fluency assessment. Completion within one sitting is recommended but not required.

Section 5: Scoring and Reporting

This chapter summarizes the scoring and reporting processes for MAP Reading Fluency. All student responses are scored automatically. Oral Reading measures that yield scaled words correct per minute (SWCPM) scores are scored by the LanguaMetrics software embedded in the test engine. All other measures are selected-response and are scored dichotomously, either correct or incorrect, at the item level by the test engine. In addition to the raw scores, one of the following performance levels is assigned to the results in each domain: *Exceeds Expectation*, *Meets Expectation*, *Approaching Expectation*, and *Below Expectation*. MAP Reading Fluency also reports a student score from the Lexile® Framework for Oral Reading (MetaMetrics, 2021). The Individual Student Report shows all raw scores achieved on a given assessment. A summary of student performance across all oral passage reading attempts is provided across three dimensions: oral reading rate, decoding accuracy, and passage comprehension. Profile statements are generated for each test and are linked to suggested instructional next steps.

Section 6: Technical Characteristics

This section presents technical information on the measurement characteristics of the MAP Reading Fluency's Foundational Skills and Oral Reading Fluency scores, including reliability and validity evidence. Data for all newly conducted analyses were collected during the 2020–2021 school year. From Fall 2020 through Spring 2021, 407,964 students received Foundational Skills scores and 277,920 students received Oral Reading Fluency scores.

Section 7: Dyslexia Screener

This chapter describes the MAP Reading Fluency Dyslexia Screener that assesses key reading skills, including those most often associated with dyslexia, without the need for a separate assessment. A predictive model flags student results that suggest possible risk factors for dyslexia or other reading difficulties. The Sentence Reading Fluency score and domain scores in Phonological Awareness, Phonics & Word Recognition, and Language Comprehension are included in a multivariate predictive model flagging students at risk of dyslexia. RAN scores supplement the multivariate predictive model. Dyslexia screening data support improved outcomes in three broad ways: (1) students *flagged* as at increased risk are flagged by a multivariate predictive model, (2) MAP Reading Fluency reports support greater *data-based differentiation* for all students, and (3) *instructional time* is returned via the efficiency of automatic, adaptive, group-administered screening of all students. A flag on these reports is not a diagnosis of dyslexia or of a reading disability; rather, it is an indicator that the student's performance suggests possible risk factors for dyslexia or some other reading difficulty.

1. Introduction

English MAP® Reading Fluency™ is an adaptive online assessment of early literacy that supports students on their path to reading comprehension by assessing and helping to improve both oral reading fluency and foundational reading skills. It is designed for students who do not yet read with solid fluency and understanding and adapts to accommodate pre-readers, early readers, and independent readers in pre-K to Grade 5 with the goal of helping all students be able to read with comprehension. The test adapts in real-time and presents easier or more difficult passages and items depending on student performance. All spoken-response tasks are scored via automated speech recognition (ASR) software. English MAP Reading Fluency also provides scores for universal screening for reading difficulties, progress monitoring, and dyslexia screening. MAP Reading Fluency is available in both English and Spanish.[1]

1.1. English MAP Reading Fluency Overview

Students generally take a 20-minute benchmark assessment three times a year in fall, winter, and spring that is automatically scored and generates actionable data about their reading skills and instructional needs, including a universal screening outcome. Alternative forms are also available, such as the dyslexia screener that offers a dyslexia screening outcome and an abbreviated progress monitoring form to assess students more frequently who are receiving intervention in reading. Any MAP Reading Fluency assessment can be administered in a group setting rather than one-on-one, which saves teachers valuable classroom instructional time. Administration procedures can vary to accommodate a variety of student and educator needs. In general, students wear headsets with microphones and read the test content out loud into the microphone. The audio is recorded and scored automatically by the speech scoring engine rather than relying on human judgment and individual administration.

While teachers can choose from various test forms, the Adaptive Oral Reading form is the default. To start, a narrator greets the students and confirms that they understand the directions. Each student reads a picture-supported story in either a picture book or graphic novel format to get started. They then read sentences silently and identify a matching picture, which gauges if the student is ready to read passages. If so, they read up to three passages out loud (although they are evaluated on only two). After reading, students answer selected-response items to demonstrate their comprehension. If the student is not ready to read passages, they are presented with a series of measures that assess foundational reading skills, including phonological awareness, early phonics and word recognition, listening comprehension, and picture vocabulary. The assessment is automatically scored, with results appearing in the educator reporting site. The following results are provided for each student: proficiency relative to grade-level expectations, individualized literacy profile, and recommended next steps.

For students who are not ready for passages, the Foundational Skills report shows their proficiency in decoding skills and oral language comprehension. Proficiencies in Phonological Awareness and Phonics & Word Recognition are each reported in the context of a learning progression. Student performance on these early literacy skills is compared to grade-level expectations and reported as *Exceeds Expectation*, *Meets Expectation*, *Approaching Expectation*, or *Below Expectation*.

[1] Details of the Spanish MAP Reading Fluency test design and content are available in a separate report (NWEA, 2021a).

For students who read the passages, the Oral Reading Fluency report shows the scaled words correct per minute (SWCPM), decoding accuracy, and passage comprehension scores. Teachers can also play back the audio recording for further evaluation. Student oral reading fluency performance is compared to grade-level expectations and is reported as *Exceeds Expectation*, *Meets Expectation*, *Approaching Expectation*, or *Below Expectation*.

1.2. Background

English MAP Reading Fluency was piloted in 2016–2017, with an early adopter program released in 2017–2018. It became available to the general public in 2018–2019. The general release of Spanish MAP Reading Fluency followed shortly thereafter, with scores becoming operational in Fall 2019. Progress monitoring for oral reading fluency was introduced to English MAP Reading Fluency in Fall 2019, and the dyslexia screener was first launched for beta users in March 2021 and available to all users in Fall 2021.

Development of MAP Reading Fluency began with the desire to help all students be able to read challenging texts with excellent comprehension, which is a primary goal of early literacy instruction. In pre-K to Grade 5, research demonstrates that development of foundational reading skills supports reading fluency, which is necessary for reading with comprehension (Shanahan & Lonigan, 2010; Jenkins et al., 2003). MAP Reading Fluency focuses on early literacy skills, including foundational skills and the development of strong oral reading fluency.

The key foundational reading skill is automatic word reading. In an alphabetic language, this begins with the ability to map written letters and letter patterns to the sounds they make (i.e., decoding) (Ehri, 2005). Development of strong word decoding is supported by the precursor skills of alphabet knowledge and phonological awareness. It is moderately to strongly supported by both print concepts and oral language comprehension (Shanahan & Lonigan, 2010). As students begin to read connected text, these skills work together. In Gough and Tunmer's Simple View of Reading model (1986), this relationship is captured by the idea that passage comprehension is the product of decoding and language comprehension. As students move to reading connected text, gauging automaticity with word reading becomes an element of oral reading fluency.

Oral reading fluency assessment has become largely ubiquitous in U.S. primary grades, with many schools using a one-minute reading sample from grade-level text, scored as words correct per minute (WCPM). This approach has a substantial research base showing its value for screening and indicating growth for students at risk of underachievement in reading (Wayman et al., 2007; Jenkins et al., 2007). Particularly among students still building their passage comprehension skills, changes in oral reading fluency offer a valuable indicator of overall growth in reading proficiency (Fuchs et al., 2001). However, research has also shown that accuracy scores are useful in instructional decisions, but that this use is lost when they are subsumed into the WCPM score alone (Valencia et al., 2010; García & Cain, 2014). Researchers have long warned about possible instructional implications of assessing WCPM without comprehension (e.g., Deno, 1985). For example, some pointed to evidence that educators were beginning to equate faster oral reading with better reading (Newman, 2009; Deeney, 2010).

Many researchers assert that the construct of oral reading fluency includes prosody (i.e., a student's phrasing and expression in support of meaning) (Kuhn et al., 2010; Rasinski et al., 2011; Samuels, 2006). In this case, faster reading can even be at odds with better, more prosodic reading (Daane et al., 2005; Paige et al., 2014). Still, reading that has sufficient rate, accuracy, and prosody is not the end goal. The real goal is improving comprehension of text, which is harder when either the text or the comprehension task is more complex, per contemporary models (e.g., RAND Reading Study Group, 2002; Common Core State Standards Initiative, 2010). With oral reading, complex comprehension tasks are typically a poor fit since they often require revisiting the text for analysis. Instead, raising the text complexity offers a way to gauge growth in reading with comprehension.

Critics of one-minute WCPM measures argue that a more robust approach to assessing oral reading fluency allows students to read a complete passage aloud and then answer comprehension questions about it (Samuels, 2007; Lipson & Wixson, 2012). From such an administration, Valencia et al. (2010) provide evidence that four types of scores each contribute to a best prediction of general passage comprehension: rate, accuracy, prosody, and comprehension. These four data points, they argue, are also those that best enable individualized instruction. MAP Reading Fluency is modeled by this more robust approach.

1.3. Design Rationale

The adaptive and group-administered approach of MAP Reading Fluency is designed to relieve teachers of lengthy assessment procedures so they can maximize instructional time. It is also designed to help tailor instruction to students' needs through effective data-based differentiation. Specifically, one purpose of MAP Reading Fluency is to point foundational skills and oral reading fluency data at immediate instructional decisions such as finding appropriate instructional emphases for sets of students; gauging the need for scaffolding and support in classroom-wide, grade-level instruction; and screening for students who are most likely to benefit from allocation of additional instructional resources. When instructional resources are allocated in the context of tailored instruction, intended outcomes are supported. When all students have strong foundational skills, fluency is supported; when all students have strong reading fluency, reading with comprehension is supported. MAP Reading Fluency results are designed to achieve these outcomes.

The goal of MAP Reading Fluency is to bring rich information from oral reading, automatically scored, to the task of individualizing reading instruction. MAP Reading Fluency is also designed to offer one source of data for comparing a student's reading fluency to a general grade-level expectation. For example, when a student's SWCPM score falls below the 25th percentile on published national norms (Hasbrouck & Tindal, 2017), reports recommend increased focus and intensity of instruction. While all students reading with fluency is the direct goal of MAP Reading Fluency, the design decision was to gauge growth in the foundational skills that support future reading fluency for students not yet able to read passages. With this in mind, MAP Reading Fluency accomplishes, but is not limited to, the following: (1) gauges student readiness for oral reading from passages, (2) informs instruction for students who cannot yet read passages, and (3) assesses oral reading proficiency and improvement.

1.3.1. Gauging Student Readiness for Oral Reading from Passages

Reading a sentence silently with sufficient speed, accuracy, and literal comprehension indicates a level of proficiency with connected text that word reading alone cannot. In MAP Reading Fluency, silent sentence fluency measures are presented to all students to help discern possible readiness for oral passage reading. Research supports the value of a measure wherein students read isolated sentences quickly and silently, then mark a quick semantic judgement. Examples include the Woodcock Johnson's Reading Fluency Task (Schrank et al., 2004) and the Test of Silent Reading Efficiency and Comprehension (Wagner et al., 2010). Stronger readers' comprehension is highly correlated to sentence-level silent fluency: students who do well on silent sentence fluency are likely to read with good phrasing when reading aloud (Klauda & Guthrie, 2008). While word reading is a stronger predictor of passage comprehension for weaker readers, silent sentence reading fluency has a tighter relationship to comprehension for stronger readers (Kim et al., 2011).

1.3.2. Informing Instruction for Students who Cannot Yet Read Passages

Consider a student who reads 18 WCPM and is at an exciting beginning point in learning to read connected text. However, this student's reading fluency is not at a point where they would be expected to understand what was read. In fact, reading more than a sentence at a time still presents a significant challenge. For a student at this level, reading aloud from passages is not a best use of time for informing instruction. Valuable information for instruction for these early readers comes from data on two broad components that feed future reading with comprehension: (1) foundational decoding skills and (2) language comprehension.

Decoding refers to phonological awareness, early phonics, and word recognition. Language comprehension refers to receptive oral vocabulary and sentence level oral language comprehension. Some students have enough language comprehension that the appropriate instructional emphasis is decoding, while others may need more emphasis on language development. Even within these broad categories, students will differ. For some students, challenges with phonemic awareness hold back word reading. For others, vocabulary may be sufficient but syntax at the sentence level can still introduce confusion.

For students who are not ready to read aloud from passages, MAP Reading Fluency collects data more useful to instruction to provide a profile of the student's foundational decoding and language comprehension skills. For example, two critical Foundational Skills domains (i.e., Phonological Awareness and Phonics & Word Recognition) each offer within-domain adaptivity. This allows the reported data to point toward a zone of proximal development (ZPD) level within a progression of skills within the domain and to offer instructional resources tightly aligned to this level. Each step in the Phonological Awareness and Phonics & Word Recognition progressions is mapped to best practice instructional materials made available by the Florida Center for Reading Research.

1.3.3. Gauging Improvements in Oral Reading

When students get better at reading texts, they improve their oral reading rate, accuracy, prosody, and passage comprehension. Often, meaningful growth is not best captured by increases in rate on the same level of material. It is unfortunate when a student who reads 130 WCPM is compelled to read faster to demonstrate growth. If students focus on reading quickly, they jeopardize their ability to make meaning from the text. When students can read passages well at a given level (i.e., showing sufficient rate, accuracy, and comprehension), faster reading does not necessarily correlate with better reading. Instead, better reading means becoming successful with harder texts and/or deeper comprehension. In MAP Reading Fluency, a student who understands what they read aloud is challenged to read from passages at a higher level of text complexity.

Students' correct words per minute are reported in terms of performance on a reference passage. In other words, passage scores are equated. Accuracy and low-level comprehension are also scored automatically, and prosody is rated by a teacher where of interest using audio playback. MAP Reading Fluency also adjusts the level of text complexity across multiple passages presented, adapting based on comprehension to find a maximum text level at which a student is showing understanding of what they read. A Lexile® oral reading measure is also reported that offers a metric for overall improvements in reading fluency, capturing together three factors: student reading rate, student reading accuracy, and the level of oral readability of the text. A rise in any of these three factors constitutes meaningful growth in oral reading fluency and will be captured as an increase in the student score using this scale.[2]

MAP Reading Fluency offers data on a student's decoding accuracy and comprehension alongside their oral reading rate to generate an individualized reader profile of strengths and needs in oral passage reading. Some students read at a fast rate but with poor accuracy on word decoding, while others read slowly and accurately. In each case, students may be successful at understanding the passage read, or they may fall short. For some students who struggle, comprehension, not decoding, is the challenge.

1.3.4. Universal Screening and Progress Monitoring

Universal screening and progress monitoring are components of a schoolwide model of student support often referred to as response to intervention (RTI) or multi-tiered systems of support. Universal screening is the component in such an approach that helps to identify students whose performance indicates some risk of poor reading outcomes (Jenkins et al., 2007). In order to best allocate increased intensity of instruction and ongoing assessment to those students most in need, data from universal screening is essential to decision making.

Progress monitoring offers an ongoing source of feedback on how students are responding to any intervention, allowing data-based adjustments to the interventions provided to students. To be meaningful, progress monitoring measures must tap a general outcome of interest (e.g., general reading proficiency) reliably and validly (Fuchs, 2004). Because of the rich and consistent body of research supporting oral reading fluency data's correlation to general reading performance and growth and because of its sensitivity to growth for progress monitoring (Wayman et al., 2007), the MAP Reading Fluency progress monitoring measure was designed using SWCPM from passage reading.

---

[2] Additional technical information about the Lexile® Oral Reading Framework is available from MetaMetrics at https://metametricsinc.com/the-lexile-framework-for-oral-reading/.

1.4. Theory of Action

Test developers posit intended interpretations and uses of their test scores and desired outcomes for their testing programs. A theory of action makes such interpretations, score uses, outcomes, and the relationships among them explicit. As such, a theory of action designs in reverse: start with the intended outcomes and interpretation and work backwards step-by-step toward the design of the assessment system. The English MAP Reading Fluency theory of action shows the hypothesized mechanisms of change and intermediate goals leading to the overarching goal of helping all students read fluently with comprehension (Jiban & Simpson, 2021). You can view it online at https://www.nwea.org/uploads/2021/06/MAP-Reading-Fluency-Theory-of-Action-Infographic-AUG21.pdf.

2. Test Design

The MAP Reading Fluency test design is based on the Simple View of Reading model (Gough & Tunmer, 1986), a research-validated model of reading development that proposes that two broad factors enable or limit comprehension: decoding and language comprehension. MAP Reading Fluency was developed to assess oral reading fluency, as well as the foundational skills in both decoding and language comprehension that lead to reading fluency. In English, when decoding is weak, even a student with excellent oral language comprehension cannot fully comprehend the text.

While the English MAP Reading Fluency assessments are aligned to different state standards, this technical report focuses on the Common Core State Standards (CCSS; National Governors Association Center for Best Practices & Council of Chief State School Officers [CCSSO], 2010). English MAP Reading Fluency aligns all the decoding, language comprehension, and fluency measures to the CCSS. In the CCSS, the foundational skills strand includes decoding and fluency components, while language comprehension skills are distributed in other strands. Each measure's alignment to the CCSS is presented in this report.

2.1. Domains and Measures

MAP Reading Fluency offers two broad sets of content: oral reading fluency and assessment of foundational reading skills. In the fully adaptive form, students are routed to one or the other of these sets of content based on their performance on a Sentence Reading Fluency measure. Students who pass the measure continue to the Oral Reading Fluency track with passages, whereas those who do not pass the measure continue to Foundational Skills. Each of these broad sets of content can also be assigned directly by the teacher.

Table 2.1 presents the domains and measures that comprise MAP Reading Fluency. The Foundational Skills track includes all measures in the Phonological Awareness, Phonics & Word Recognition, and Language Comprehension domains. Phonological Awareness and Phonics & Word Recognition each measure knowledge and skills pertaining to decoding (i.e., the process of mapping print to sound). Measures in the decoding domains are administered adaptively. The measures in the table range from the lowest zone of proximal development (ZPD) level (i.e., the first developing skills) to the highest, encompassing single letter or letter sound up to word reading and phoneme manipulation.

The decoding measures are speeded so that gains in proficiency are captured both by accurate responding and by rate of responding. Students typically move from accurate but slower responding to a faster rate of responding, which indicates increasing automaticity with the skill. Students see as many items as their rate allows in the allotted time of either one or two minutes depending on the measure. In the Language Comprehension domain, gains in proficiency are gauged by increased accuracy with the vocabulary and listening comprehension measures. In this domain, accuracy is relevant but rate is not, so the measures use a fixed number of items rather than a fixed time duration. For students at beginner levels of foundational skills, the Print Concepts content is administered as an additional domain. These measures of understanding about print are also not speeded. For students who are routed (or directly assigned) to the Oral Reading Fluency test content, passages are presented for reading aloud and followed by six non-speeded basic comprehension questions.

Termination of a speeded measure is based on a fixed duration (i.e., one or two minutes) rather than on a fixed number of items. For all non-speeded measures, students taking longer are prompted to respond to the item in the interest of not letting students stall or take too much time on a single item. After the prompt, students who have not responded after a maximum amount of time are moved to the next item. Passages do not have a prompt; rather, if students take longer than five minutes to read a passage aloud, the screen fades out and the student is moved on to the next passage or item.

Table 2.1. Assessed Domains and Measures of English MAP Reading Fluency

Domain	Measure	Code	Duration
Foundational Skills			
Phonological Awareness	Rhyme Completion	030	2 minutes, speeded
	Counting Syllables	017	1 minute, speeded
	Onset-Rime Blending	018	1 minute, speeded
	Initial Sound Matching	001	2 minutes, speeded
	Blending Phonemes	019	1 minute, speeded
	Phoneme Counting	020	1 minute, speeded
	Phoneme Addition/Deletion	021	2 minutes, speeded
	Phoneme Substitution	022	2 minutes, speeded
Phonics & Word Recognition*	Letter Knowledge	002	1 minute, speeded
	Letter-Sound Fluency	003	1 minute, speeded
	Build Words: One Letter	024	1 minute, speeded
	Word Families: Initial Letter	023	1 minute, speeded
	Decoding: CVC	007	1 minute, speeded
	Build Words: CVC	025	2 minutes, speeded
	Decoding: Single Syllable	027	1 minute, speeded
	Build Words: Single Syllable	026	2 minutes, speeded
	Sentence Reading Fluency	008	2 minutes, speeded
Language Comprehension	Picture Vocabulary	005	15 items, up to 30 seconds per item
	Listening Comprehension	004	15 items, up to 30 seconds per item
Print Concepts	Print Concepts	031–036	6 items, up to 45 seconds per item
Oral Reading Fluency			
Oral Reading**	Oral Reading: Picture Book/Graphic Novel	013/040	Up to 5 minutes
	Oral Reading: Passages	011	Up to 5 minutes per passage
	Oral Reading: Passage Comprehension Quiz	014	6 items, up to 90 seconds per item

*The Sentence Reading Fluency measure is not used in the students' Phonics & Word Recognition score.
**Oral Reading: Passages and Oral Reading: Passage Comprehension Quiz are administered as a set (i.e., students read a passage then answer items about it). Oral Reading: Passages presents three passages for the benchmark forms and one passage in the progress monitoring form.

2.2. Test Forms

As shown in Table 2.2, MAP Reading Fluency provides different forms to meet the varied needs of users. The default MAP Reading Fluency benchmark test form for pre-K to Grade 3 students is Adaptive Oral Reading, which routes students to the Oral Reading Fluency track if they are ready to read passages or to the Foundational Skills track if they are not. The default for Grades 4+ students is the Passages Only form that includes only the Oral Reading Fluency track. Other benchmark forms can be substituted or administered in addition to the default forms, including Foundational Skills, Foundational Skills–Beginner, or Passages Only. The Dyslexia Screener form may also be considered a benchmark in that it includes the content from Foundational Skills, plus content specific to dyslexia screening (i.e., rapid naming). These various forms largely draw on the same operational item pool within a measure. The Progress Monitoring form is designed for more frequent administration to students receiving intervention in reading. In oral reading, each Progress Monitoring form has one passage with six questions.

On the passage track of the Adaptive Oral Reading form and on the Passages Only form, students receive up to three passages, each with six comprehension items. The third passage, if administered, is a field test passage. The passage content is designed for the student's rostered grade level, with the initial passage being on-grade level and next passages adapting to harder or easier text based on student comprehension performance.

Table 2.2. MAP Reading Fluency Test Forms

Form	Description	Languages	Adaptive?	Universal Screener?
Adaptive Oral Reading	Assigned by default to students in pre-K to Grade 3. Directs students into either oral reading fluency and comprehension or foundational skills, depending on each student's performance on Sentence Reading Fluency. Content is presented according to adaptive test logic based on student performance within the test session.	English and Spanish	Yes	Yes
Foundational Skills	Assesses Phonological Awareness, Phonics & Word Recognition, and Language Comprehension. This form does not route any students into oral reading passages.	English and Spanish	Yes	Yes
Foundational Skills–Beginner	An entry-level form intended for fall testing of Pre-K students and kindergarteners who have not been to Pre-K. It includes the full Language Comprehension domain but only the first four measures in the skill progressions within the decoding Phonological Awareness and Phonics & Word Recognition domains. It also assesses Print Concepts, including word concepts and text directionality.	English and Spanish	Yes	No

Form	Description	Languages	Adaptive?	Universal Screener?
Adaptive Oral Reading–Passages Only	Assigned by default to students in Grades 4+. Contains reading passages and comprehension questions and does not measure any foundational skills. This form is an option for students who can read connected text and for students who have tested into the oral reading pathway on previous administrations. Picture Book/Graphic Novel and Sentence Reading Fluency are included in this format, but there is no sentence reading threshold score required to move on to passage reading.	English and Spanish	Yes	Yes (English only)
Dyslexia Screener	Includes measures from the Foundational Skills domains of Phonological Awareness, Phonics & Word Recognition, and Language Comprehension, along with Sentence Reading Fluency. Rapid Automatized Naming (RAN) data are available for students taking this form. For students who struggle with foundational reading skills, low naming speed may be an additional risk factor for difficulties in developing reading fluency.	English only	Yes	No
Progress Monitoring	Short tests designed to measure reading progress. Because students at risk in Grade 1 are typically not ready for passages for most of the school year, the earliest recommended use for monitoring students at risk is spring of Grade 1. Depending on local district policy and programming, students flagged at-risk may be enrolled in Tier 2 or Tier 3 instructional groupings. MAP Reading Fluency Progress Monitoring is appropriate for Tier 2 and Tier 3 students with oral reading fluency goals	English only	No	No

2.2.1. Adaptive Oral Reading

In the Adaptive Oral Reading form, students are routed to either Oral Reading Fluency (the passage track) or to Foundational Skills based on their performance at the beginning of the assessment. Both formats are administered adaptively and present the same measures, with the exception of (1) the Oral Reading measures that are only on the Oral Reading Fluency forms and (2) Print Concepts that is only on the Foundational Skills–Beginner form.

To start the assessment, all students read a picture book or graphic novel and complete the two-minute Sentence Reading Fluency measure, the routing test for all Adaptive Oral Reading forms. A threshold raw score (15 or more) and accuracy rate (75% or more) for Sentence Reading Fluency must be obtained to proceed to Oral Reading: Passages and Comprehension Quiz, with the exception of students in fourth grade and above who will always proceed to Oral Reading: Passages and Comprehension Quiz. Students performing below this threshold are presented instead with decoding and language comprehension measures in the Foundational Skills track.

Students routed to the passage track receive up to three passages, each with six comprehension items. For students in pre-kindergarten to first grade routed to the passage track, a third field test passage is not administered if the student did not pass the basic comprehension quiz for either of the first two passages (≤ 66% correct). In that case, language comprehension measures are administered instead. No student takes all of the Foundational Skills measures. Each Foundational Skills test event includes a subset of measures in Phonological Awareness and Phonics & Word Recognition, selected adaptively based on performance within a progression of skills.

2.2.2. Dyslexia Screener

The MAP Reading Fluency Dyslexia Screener allows teachers to assess an entire in as little as 20 minutes. It assesses key reading skills, including those most often associated with dyslexia, without the need for a separate assessment. A predictive model flags student results that suggest possible risk factors for dyslexia or other reading difficulties. It was first launched to beta users in March 2021 and was to all users in Fall 2021. See Section 7 for more details.

2.2.3. Progress Monitoring

Progress monitoring provides a quick and reliable way to measure improvement in reading over time. The progress monitoring test is 5–10 minutes in length and currently available for English Oral Reading Fluency only. Progress monitoring tests are not adaptive. Once a progress monitoring test has been assigned to a student, they will be presented with a new passage at their assigned Lexile® level every time they log in. This will continue until the proctor stops progress monitoring. If a benchmark test is assigned to a student who also has progress monitoring assigned, the benchmark test will be presented the next time the student logs in and, once it is done, the system will go back to presenting progress monitoring tests the next time the student logs in. When students take a progress monitoring test, they are presented with one passage that they read out loud, followed by six questions. The passages are drawn from a bank, so students see different passages each time they test. Passages repeat after the entire bank of passages at the assigned Lexile® level has been presented once. If progress monitoring is used, it is recommended to start after first administering a benchmark test to determine a student's reading level. However, progress monitoring can be assigned at any time.

2.3. Phonological Awareness

Early learners' phonemic awareness is among the strongest predictors of future decoding proficiency in English (Gillon, 2004; Melby-Lervåg et al., 2012). The skills children use in working with larger units of sound and eventually individual phonemes feed their growing ability to decode unfamiliar words by sounding words out (Adams, 1990). Research has converged on a general sequence of development in phonological awareness, one that holds true across languages even as its rapidity is influenced by linguistic and educational contexts (Anthony & Francis, 2005). The sequence moves from large units of sound, such as words, to smallest units of sound, or phonemes.

In English, children develop sensitivity to *whole words* as sounds before parts of words such as syllables. Next, they hear and work with *parts of syllables* such as onsets and rimes. Finally, children develop the ability to distinguish and work with individual phonemes. For any unit of sound, blending typically develops before segmenting (Anthony & Francis, 2005). Last to fully develop is the ability to manipulate phonemes, including phoneme addition, deletion, and substitution (Anthony & Francis, 2005; Moats & Tolman, 2009; Gillon, 2017).

Strength at the level of manipulating individual phonemes appears to be the most closely correlated to word decoding in English (Kilpatrick, 2012b; Melby-Lervåg et al., 2012). As children move beyond accuracy to automaticity on these skills, this more automatic "phonemic proficiency" enables orthographic mapping, or assigning a spelling to each sound in a word that a student has read (Kilpatrick, 2018). Orthographic mapping is how readers move a word into memory so that the word becomes part of the reader's set of instantly recognized sight word lexicon (Ehri, 2014). Because the automatic nature of phonemic proficiency matters for word recognition, speededness is an important element in assessing phonological awareness.

It is useful to find students *earlier* who are not on track toward phonemic awareness or proficiency. For early screening of students at risk of later reading failure, measures of earlier-developing phonological awareness skills have proven valuable (O'Connor & Jenkins, 1999). The Phonological Awareness measures are designed to fit this research-based progression, with two measures at each of four levels as shown in Table 2.3.

Table 2.3. Phonological Awareness Progression

Level 1: Rhymes and Syllables	Level 2: Initial Sounds	Level 3: Blending Phonemes and Segmenting	Level 4: Phoneme Manipulation
Rhyme Completion *Measures phonological rhyme identification skills*	**Onset-Rime Blending** *Measures initial phoneme blending skills*	**Blending Phonemes** *Measures phoneme blending skills*	**Phoneme Addition/Deletion** *Measures phoneme manipulation skills*
Counting Syllables *Measures phonological syllable segmenting skills*	**Initial Sound Matching** *Measures initial phoneme identification skills*	**Phoneme Counting** *Measures phoneme segmenting skills*	**Phoneme Substitution** *Measures phoneme manipulation skills*

2.3.1. Level 1: Rhymes and Syllables

At the earliest stages of phonological awareness, children are still developing the ability to distinguish between whole words and syllables. Mesmer & Williams (2015) found that until children have good awareness of syllables, mastery of the concept of "word" remains precarious. After children can blend syllables, they begin to work with segmenting them within words. Children who can clap out or count the syllables in a word are demonstrating their ability to segment (Gillon, 2004). Sensitivity to rhyming develops early in the progression of phonological awareness as well (Moats & Tolman, 2009). One-syllable rhyming words differ in their onset but have a shared rime. Hearing rhyming words is therefore a step toward working with onset-rime blending and segmentation. Rhyme sensitivity strongly predicts later development of phonemic awareness skills (Anthony & Francis, 2005).

Table 2.4. Specifications—Rhyme Completion

Code	030
Specifications	Students choose the third word completing a trio of rhyming words, where the first two rhyming words are given. Replayable audio gives the names of the four onscreen pictures. No text is onscreen. Words included in the measure are required to be one-syllable words commonly familiar to kindergarten students. Any that were not clearly depictable by a simple illustration are rejected. Score is correct selections over 2 minutes.
Item Pool	Up to 30 items presented in random order
Duration	2 minutes, speeded
CCSS Alignment	K.RF.2.a – Recognize and produce rhyming words.

Figure 2.1. Sample Item—Rhyme Completion

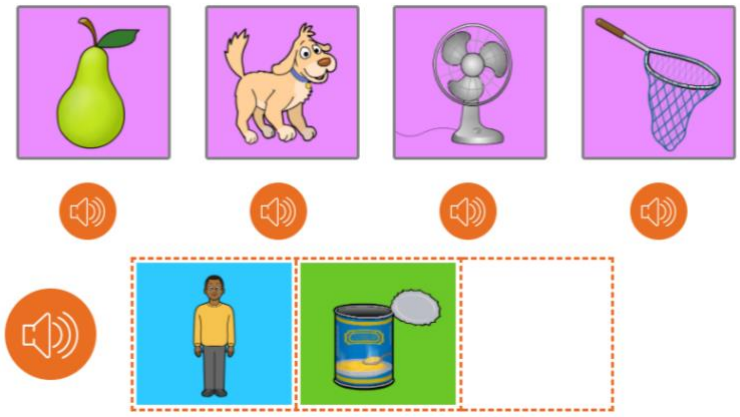
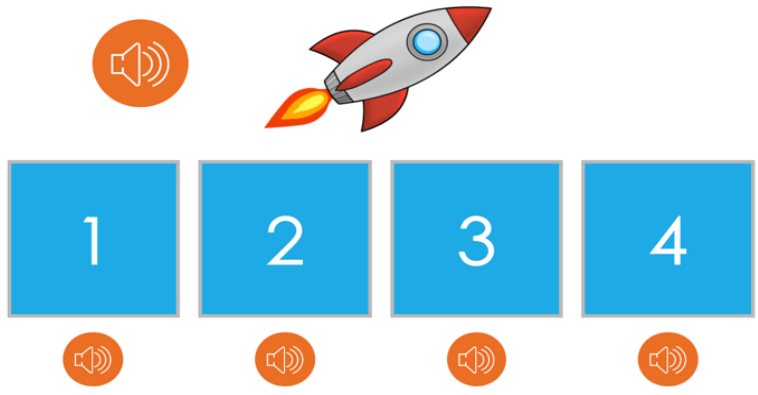
Rhyme Completion	Listen to four word choices. Given the first two words in a rhyming set, choose the word that completes the trio of rhyming words.	

Table 2.5. Specifications—Counting Syllables

Code	017
Specifications	Students choose the number of syllables in a spoken word. The word is given in audio and supported with a picture. The student then segments and counts the syllables, choosing a numeral from 1 to 4 as a response. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 29 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	K.RF.2.b – Count, pronounce, blend, and segment syllables in spoken words.

Figure 2.2. Sample Item—Counting Syllables

Counting Syllables	Listen to a word. Count the syllables and choose the number.	

2.3.2. Level 2: Initial Sounds

As children move to smaller sound units than the syllable, they begin by working with the two parts of a syllable: the onset (the initial sound or sounds before the vowel) and the remaining rime. When children can hear and work with initial sounds, they have progressed from processing larger phonological chunks (i.e., whole words or syllables) to the beginning of *phoneme* level awareness, or distinguishing single sounds. For children learning to read in English, development of *phoneme* level understanding and flexibility both supports and benefits from skills with letter sounds (Perfetti, 1997; Shanahan & Lonigan, 2010).

Blending is generally an easier task than segmenting, and it is easier to blend the onset and rime than to blend individual phonemes. While phoneme level awareness is a stronger predictor of reading proficiency, onset-rime level awareness constitutes a step toward phonemes (Cassady & Smith, 2004). Moreover, learning to blend gives children a tool they eventually use directly in decoding, especially when decoding by analogy to other words with the same rime (Goswami & Mead, 1992). In some measures requiring students to orally produce the initial sound in a word, scoring reliability has been difficult to achieve (e.g., Cummings et al., 2011). Similarly, speech scoring is not sufficiently reliable on single phoneme production in isolation. Because of this, MAP Reading Fluency assesses initial sound understanding through selected-response items.

Table 2.6. Specifications—Onset-Rime Blending

Code	018
Specifications	Students blend a given onset and rime into a word and choose the image that depicts that word. The onset and rime are given in audio, separated by a pause. Words used include only single-syllable, three phoneme words with medial vowel. All words must be clearly depictable in a simple image; a word like “his” would not meet this criterion. Distractors include at least one phoneme in common with the correct word. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 45 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	K.RF.2.c – Blend and segment onset and rime of single-syllable spoken words.

Figure 2.3. Sample Item—Onset-Rime Blending

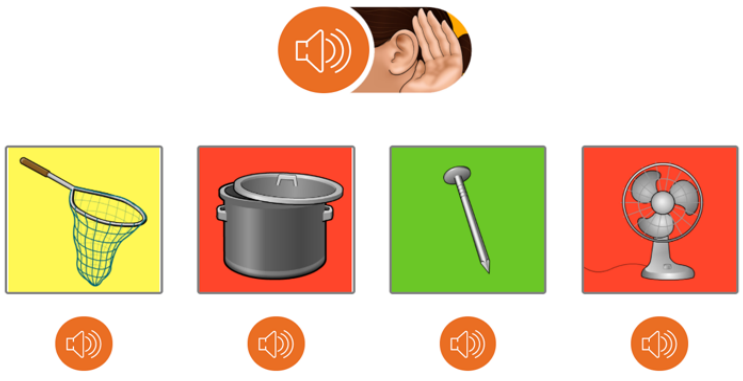
Onset-Rime Blending	Listen to an isolated initial sound and rime. Blend the sounds together and choose the word.	

Table 2.7. Specifications—Initial Sound Matching

Code	001
Specifications	Students select the two words with the same initial sound. Audio gives the names of the four onscreen pictures, each beginning with a simple consonant or digraph phoneme. No text is onscreen. Words included in the measure are required to be one-syllable words commonly familiar to kindergarten students. Any that are not clearly depictable by a simple illustration have been rejected. Score is correct pair selections over 2 minutes.
Item Pool	Up to 26 items presented in random order
Duration	2 minutes, speeded
CCSS Alignment	K.RF.2.d – Isolate and pronounce the initial, medial vowel, and final sounds (phonemes) in three-phoneme (consonant-vowel-consonant, or CVC) words.

Figure 2.4. Sample Item—Initial Sound Matching

Initial Sound Matching	Listen to four words. Choose the two with the same beginning sound.	

2.3.3. Level 3: Blending Phonemes and Segmenting

When children move from broader phonological awareness to *phonemic* awareness, they are demonstrating the skills that most directly support and predict decoding in alphabetic languages. It is at this phoneme level that sound awareness offers the strongest concurrent and longitudinal prediction of reading proficiency (Hulme et al., 2002; Melby-Lervåg et al., 2012). Moreover, instruction in phonemic awareness has demonstrated significant positive effects on later reading proficiency in English (Ehri et al., 2001).

Phonemic blending typically develops before phoneme segmenting (Moats & Tolman, 2009; Gillon, 2004; Paulson, 2004). Phoneme-level awareness is facilitated by development of letter sound knowledge (Anthony & Francis, 2005), and both have a reciprocal relationship to the development of word decoding (Perfetti et al., 1987). Where measures require students to orally produce a single phoneme, scoring reliability is challenged for human scorers (e.g., Cummings et al., 2011). Automatic speech scoring is not sufficiently reliable on phonemes in isolation either. Because of this, MAP Reading Fluency assesses phoneme segmentation through selected-response items: when students count phonemes, they demonstrate segmentation skills.

Table 2.8. Specifications—Blending Phonemes

Code	019
Specifications	Students blend a given set of three phonemes into a word and choose the image that depicts that word. The phonemes are given in audio, separated by a pause. Words used include only single-syllable, three phoneme words with medial vowel. All words must be clearly depictable in a simple image; a word like “his” would not meet this criterion. Distractors include at least one phoneme in common with the correct word. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 43 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	1.RF.2.b – Orally produce single-syllable words by blending sounds (phonemes), including consonant blends.

Figure 2.5. Sample Item—Blending Phonemes

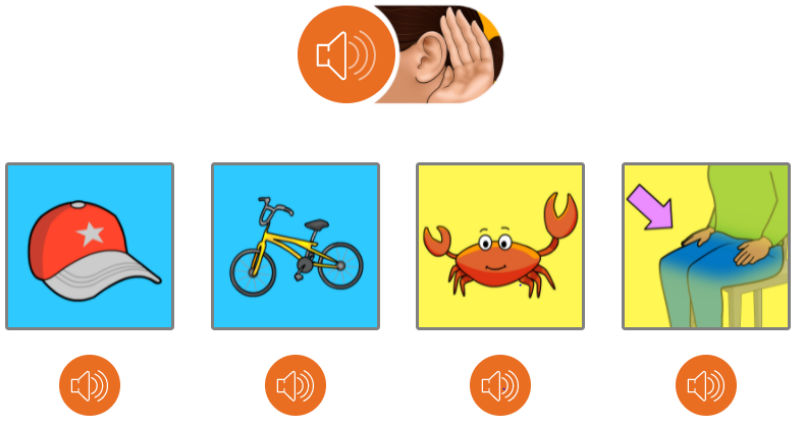
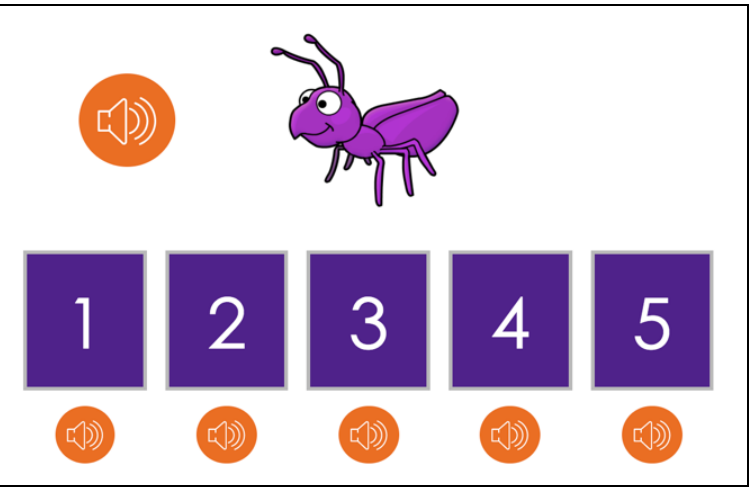
Blending Phonemes	Listen to three separated phonemes. Blend the sounds together and choose the word.	

Table 2.9. Specifications—Phoneme Counting

Code	020
Specifications	Students choose the number of phonemes in a spoken word. The word is given in audio and supported with a picture. The student then segments and counts the phonemes, choosing a numeral from 1 to 5 as a response. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 45 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	1.RF.2.d – Segment spoken single-syllable words into their complete sequence of individual sounds (phonemes).

Figure 2.6. Sample Item—Phoneme Counting

Phoneme Counting	Listen to a word aloud. Isolate the phonemes, count them and choose the number.	

2.3.4. Level 4: Phoneme Manipulation

Strong phonemic awareness goes beyond segmenting and blending phonemes. Phoneme manipulation includes some of the last-developing skills in the progression of phonological awareness: phoneme addition, phoneme deletion, and phoneme substitution (Anthony & Francis, 2005; Gillon, 2017). For children to delete or substitute a phoneme in a word, they must tap into skills in both phoneme segmentation and phoneme blending (Kilpatrick, 2012b). This flexibility with phonemes supports the decoding of unfamiliar words using analogy and sounding out strategies (Ehri, 2005). Researchers have found that tasks requiring these kinds of phoneme manipulation are among the strongest correlates of decoding proficiency in English (Catts et al., 2001; Kilpatrick, 2012a; Kroese et al., 2000; Lenchner et al., 1990). Phonemic skills at this level are developed, reciprocally, by practice with decoding words (Shanahan & Lonigan, 2010).

Table 2.10. Specifications—Phoneme Addition/Deletion

Code	021
Specifications	Students find the new word formed by adding or deleting a phoneme from a given initial word. In audio, a three- or four-phoneme word is given with an instruction about adding or deleting a particular phoneme. Each item specifies whether to add or delete the specific phoneme, as well as either the beginning or ending of the word as the location of the phoneme changes. These directions are visually supported by Elkonin boxes showing the position of the changed phoneme. Four answer options are picture words, with available audio naming the picture. No words with r-controlled or l- controlled vowels are included; no words with the letter x are included. Students form the new word mentally and then select the picture that depicts it. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 30 items presented in random order
Duration	2 minutes, speeded
CCSS Alignment	K.RF.2.e – Add or substitute individual sounds (phonemes) in simple, one-syllable words to make new words.

Figure 2.7. Sample Item—Phoneme Addition/Deletion

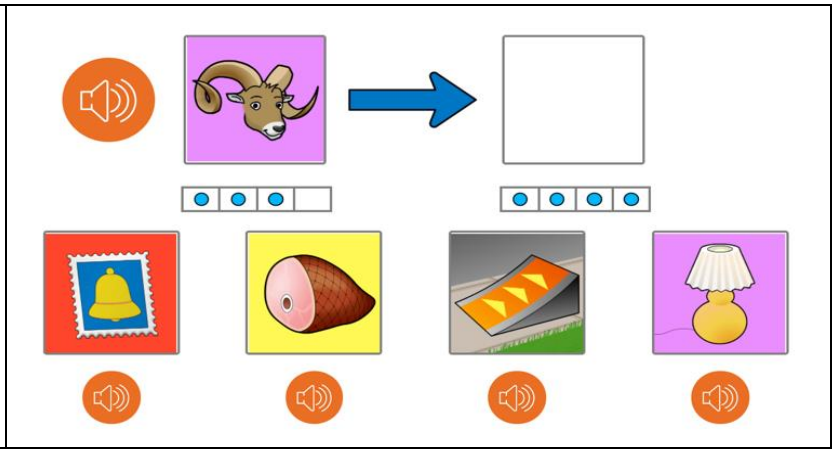
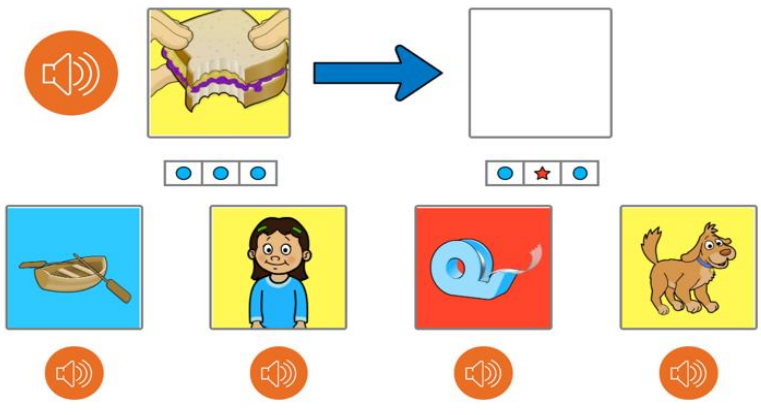
Phoneme Addition/ Deletion	Listen to a word aloud and add or subtract an initial or final sound. Choose the new word.	

Table 2.11. Specifications—Phoneme Substitution

Code	022
Specifications	Students find the new word formed by substituting a phoneme into a given initial word. In audio, a three- or four-phoneme word is given with an instruction about which particular phoneme to substitute into the word and where. These directions are visually supported by Elkonin boxes showing the position of the changed phoneme. Four answer options are picture words, with available audio naming the picture. For three phoneme, CVC style words, the medial vowel is the target of substitution. For four phoneme (CCVC, CVCC) words, the interior consonant in the consonant blend is the target of substitution. No words with r- controlled or l- controlled vowels are included; no words with the letter x are included. Students form the new word mentally and then select the picture that depicts it. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 30 items presented in random order
Duration	2 minutes, speeded
CCSS Alignment	K.RF.2.e – Add or substitute individual sounds (phonemes) in simple, one-syllable words to make new words.

Figure 2.8. Sample Item—Phoneme Substitution

Phoneme Substitution	Listen to a word aloud. Change the middle sound and choose the new word.	

2.4. Phonics & Word Recognition

Learning to decode in English is a complex undertaking. Beginning with letter sounds and moving to word reading, decoding is the task of turning sets of letters on the page into the sounds they represent. Broadly, the youngest children begin to approach word identification *logographically*, where they are in a pre-alphabetic phase: they recognize how a particular word looks without attending to letter sounds at all (Frith, 1985; Ehri, 1998). Next, after understanding the alphabetic principle, they shift to a *partial alphabetic* phase where they attend more to initial sounds in words than to medial or final sounds (Guthrie & Seifert, 1977; Ehri, 1998). Gradually, they use letter sounds and phonics patterns to move from consonant-vowel-consonant (CVC) words to single-syllable words with blends, digraphs, and long vowel spellings. Later still, they read multi-syllabic words (Guthrie & Siefert, 1977; Pirani-McGurl, 2009).

As children learn to *decode* words, they must also learn to *encode* words—to write them. After children learn letter sounds, they typically begin in a *semi-phonetic stage* of writing, characterized by use of invented spellings: they use a letter for each sound they hear in a word, sometimes skipping vowels or substituting letters as they develop their sense of the speech to print connection (Read, 1971; Gentry, 1982; Richgels, 1995). Children move from a *phonetic* stage into *correct spelling* as they gain experience with words in print (Gentry, 1982). Spelling recognition skills help predict eventual reading proficiency, even after the contributions of word reading (Katzir et al., 2006).

The Phonics & Word Recognition measures tap both decoding and encoding abilities. They are designed as a research-based progression, with two measures at each of four levels, as shown in Table 2.12.

Table 2.12. Phonics & Word Recognition Progression

Level 1: Letters and Sounds	Level 2: Letters in Words	Level 3: CVC Words	Level 4: One-Syllable Words
Letter Knowledge *Measures letter identification knowledge*	**Build Words: One Letter** *Measures letter sound decoding skills in word*	**Decoding: CVC** *Measures early word decoding skills*	**Decoding: Single Syllable** *Measures word decoding skills*
Letter-Sound Fluency *Measures letter sound correspondence knowledge*	**Word Families: Initial Letter** *Measures letter sound decoding skills in words*	**Build Words: CVC** *Measures early word encoding skills*	**Build Words: Single Syllable** *Measures word encoding skills*

2.4.1. Level 1: Letters and Sounds

A student who can name a presented letter of the alphabet quickly and accurately is likely on a better English literacy trajectory than a student who cannot (Speece et al., 2003). Because letter names are less directly applicable than letter sounds in decoding, the value of a screener using only fluency in letter naming has been questioned (e.g., Jenkins et al., 2007). However, as a proxy, letter naming offers an important window into a student's literacy experiences before schooling. The literature on screening for risk of reading failure indicates that the value of letter knowledge is strongest as one among a broader set of measures (Foorman et al., 1998; O'Connor & Jenkins, 1999).

While children may know that letters have names, the understanding that each makes a sound in reading is a separate and important step. Research evidence points to the utility of letter sound fluency in screening for risk of reading failure, both alone (Speece & Case, 2001; Speece, 2005) and in combination with other brief measures (O'Connor & Jenkins, 1999).

Table 2.13. Specifications—Letter Knowledge

Code	002
Specifications	Each item presents in audio the name of a letter, and eight uppercase letters are presented onscreen. Incorrect options include letters that bear visual resemblance to the correct letter but do not rhyme or sound similar (e.g., for letter F, the letter S is not presented as an option). Only uppercase letters are assessed to distinguish the task clearly from the Letter-Sound Fluency task that uses lowercase letters. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 20 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	K.RF.1.d – Recognize and name all upper- and lowercase letters of the alphabet.

Figure 2.9. Sample Item—Letter Knowledge

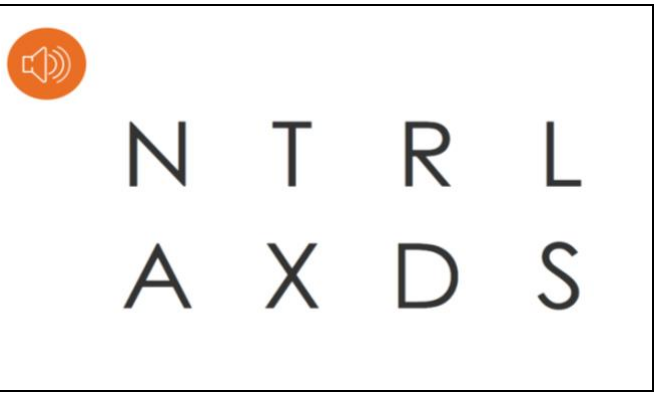
Letter Knowledge	Choose the named letter.	

Table 2.14. Specifications—Letter-Sound Fluency

Code	003
Specifications	Each item presents in audio the sound of a letter and an example word beginning with that sound (e.g., /p/, as in “party”). Eight lowercase letters are presented onscreen. Incorrect options include letters that are both close and far in terms of articulation (e.g., other stops, but also fricatives or liquids). Only lowercase letters are assessed to distinguish the task clearly from the Letter Knowledge task that uses uppercase letters. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 20 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	K.RF.3.a – Demonstrate basic knowledge of one-to-one letter-sound correspondences by producing the primary or many of the most frequent sound for each consonant. K.RF.3.b – Associate the long and short sounds with common spellings (graphemes) for the five major vowels.

Figure 2.10. Sample Item—Letter-Sound Fluency

Letter-Sound Fluency	Listen to an isolated sound and a word that starts with it. Choose the letter that makes the sound.	

2.4.2. Level 2: Letters in Words

The alphabetic principle (i.e., the realization that each letter conveys a sound *in text*, in the order in which they are presented) is the central realization upon which decoding in English rests. Children do not make this realization until they have learned to recognize some letters and name them (Adams, 1990; Ehri, 2002). Children can then work with letter sounds in the context of whole words. As they tackle words, children begin in a *partial alphabetic* phase where they use any phoneme they can distinguish but may not use all of them present in a word (Ehri, 1998). In English, children typically first attend more to initial letter sounds in words than to any other sounds, and they use final consonants more readily than medial vowels (Guthrie & Seifert, 1977; Morris et al., 2003). In English, words with the same rime (sometimes called "word families") offer an analogy-based route to early whole word decoding (Treiman et al., 1995; Walton & Walton, 2002).

Table 2.15. Specifications—Build Words: One Letter

Code	024
Specifications	Students hear a word read aloud and see an accompanying picture. The onscreen text shows the word with one letter missing. Students choose the missing letter, which pops to the word. Words in this measure are all CVC words and must be depictable enough that the audio for the word is supported by the picture for clear discernment. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 45 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	K.RF.3d – Distinguish between similarly spelled words by identifying the sounds of the letters that differ.

Figure 2.11. Sample Item—Build Words: One Letter

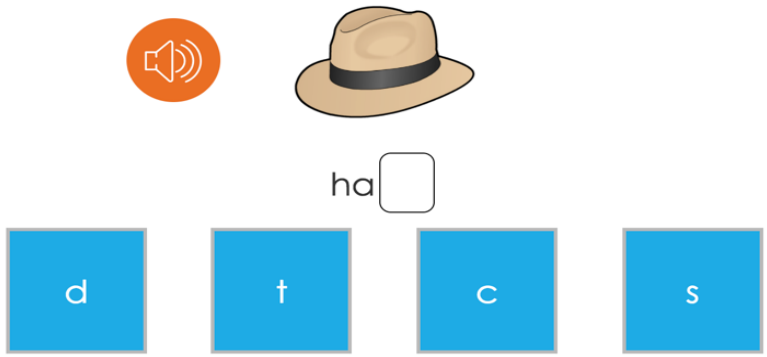
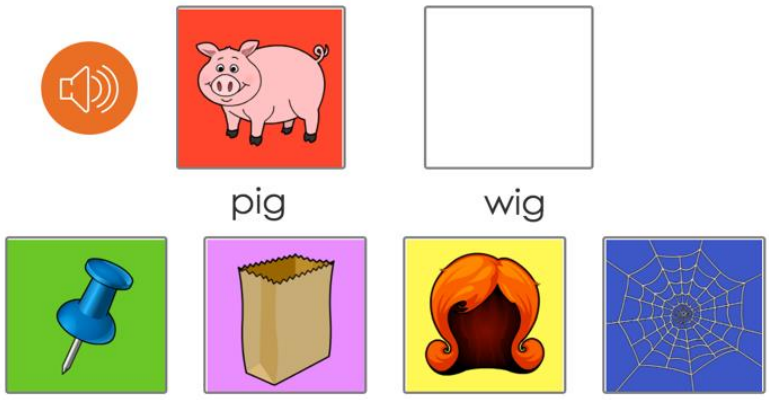
Build Words: One Letter	Listen to a word and complete its spelling by choosing a letter for the missing sound.	

Table 2.16. Specifications—Word Families: Initial Letter

Code	023
Specifications	Students hear and see an example word, supported by a picture. A second word is shown onscreen for them to read, without audio or picture. The two CVC words share a rime; they are from the same "word family" (e.g., pig and wig). The student reads the second word, perhaps by analogy to the given first word, and selects the picture that matches that second word. A next item is only presented after a selection is made. Score is correct selections per minute.
Item Pool	Up to 35 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	K.RF.3d – Distinguish between similarly spelled words by identifying the sounds of the letters that differ.

Figure 2.12. Sample Item—Word Families: Initial Letter

Word Families: Initial Letter	Look at two words from the same CVC Word Family, one paired with a picture and read aloud. Decode the second word and choose the correct picture pairing.	

2.4.3. Level 3: CVC Words

In English, words with the CVC structure are highly regular, representing three phonemes with the middle being a short vowel sound. For these words, letter sound knowledge and phoneme blending come together as word decoding (Adams, 1990). Assessments of word-level decoding fluency in English have included both word reading and "nonsense word" reading. Fuchs et al. (2004) found that real-word reading had superior concurrent validity. As children begin to decode the letter sounds in words, they also begin to encode, or write: they form their own words with letters. Snow et al. (1998) demonstrate that phonemic skills and letter knowledge collaborate to form word encoding – invented and then conventional spelling. Spelling shares much with decoding in that they map sound and print together (Robbins et al., 2010; Nunes et al., 2012).

Table 2.17. Specifications—Decoding: CVC

Code	007
Specifications	Silent measure. The task is to read the onscreen word and choose the onscreen picture that depicts the word from among four onscreen pictures total. The pool of words is composed of phonetically regular, CVC words using short vowel sounds (e.g., dog). Each word is required to be clearly depicted in a simple illustration (e.g., the word "get" does not meet this requirement). The illustration for each word in the pool appears onscreen with three other illustrations, each designed as much as possible to depict a feasible misreading of the onscreen word. For example, where the word is "cat," other illustrations might show "coat" or "can." A selection must be made for the student to go on to the next item. Score is correct selections per minute.
Item Pool	Up to 51 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	1.RF.3.b – Decode regularly spelled one-syllable words.

Figure 2.13. Sample Item—Decoding: CVC

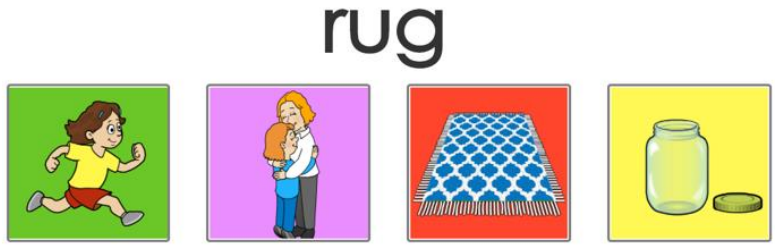
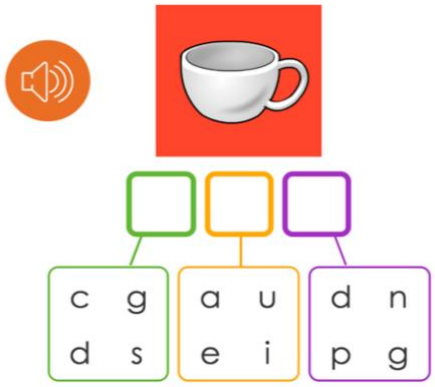
Decoding: CVC	Decode the onscreen word and choose the picture that matches.	

Table 2.18. Specifications—Build Words: CVC

Code	025
Specifications	Students build a given word using a set of letter options for each position in the word. The CVC word is given in audio and shown in a picture, and three empty boxes are shown in which students will pop one letter apiece to spell the word. A set of four consonants is given as answer options for the first box, four vowels are given for the second box, and four consonants are given for the third box. Score is correct box completions per minute.
Item Pool	Up to 45 items presented in random order, each with three scorable boxes
Duration	2 minutes, speeded
CCSS Alignment	1.RF.3 – Know and apply grade-level phonics and word analysis skills in decoding words. 1.RF.3.b – Decode regularly spelled one-syllable words.

Figure 2.14. Sample Item—Build Words: CVC

Build Words: CVC	Listen to a word and complete its spelling by choosing a letter for each sound.	

2.4.4. Level 4: One-Syllable Words

Typically, after children can read and build words in English with the CVC structure, they develop skill with words of other definable structures such as CCVC, CVCC, CCVCC, and CVCe (with silent final -e). Only slightly harder are single-syllable words with vowel combinations, including long vowel sounds and other sounds like -oo- and -oi- (Guthrie & Seifert, 1977; Pirani-McGurl, 2009). Lists of words with a variety of these regular grapho-phonemic patterns have been used in timed word reading fluency measures. Compared with other brief screening measures designed to flag first graders at risk of poor reading outcomes, word identification fluency is among the strongest (Clemens et al., 2011).

In addition to decoding, *encoding* of various single-syllable words relies on grapho-phonemic knowledge—not just individual letter sounds, but also larger units such as vowel combinations and consonant digraphs (Robbins et al., 2010; Nunes et al., 2012). Identifying correct spelling patterns in English matters: Katzir et al. (2006) found that spelling recognition explained significant variance in passage comprehension, even after the effects of word reading proficiency had been included.

Table 2.19. Specifications—Decoding: Single-Syllable

Code	027
Specifications	Silent measure. The task is to read the onscreen word and choose the onscreen picture that depicts the word from among four onscreen pictures total. The pool of words is composed of one-syllable words that are all phonetically regular, following systematic phonics rules. Words include long vowels using vowel pairs or final silent *e* (e.g., boat or vote), additional vowel variants (e.g., coin, crown), initial or final digraphs (e.g., chop or sing), and initial and final consonant blends (e.g., stop). Each word must be clearly depicted in a simple illustration. For example, the word "that" does not meet this requirement. The illustration for each word in the pool appears onscreen with three other illustrations, each designed as much as possible to depict a feasible misreading of the onscreen word. For example, where the word is "coat," other illustrations might show "cat" or "cot." A selection must be made for the student to go on to the next item. Score is correct selections per minute.
Item Pool	Up to 30 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	1.RF.3.b – Decode regularly spelled one-syllable words. 1.RF.3 and 2.RF.3 – Know and apply grade-level phonics and word analysis skills in decoding words.

Figure 2.15. Sample Item—Decoding: Single-Syllable

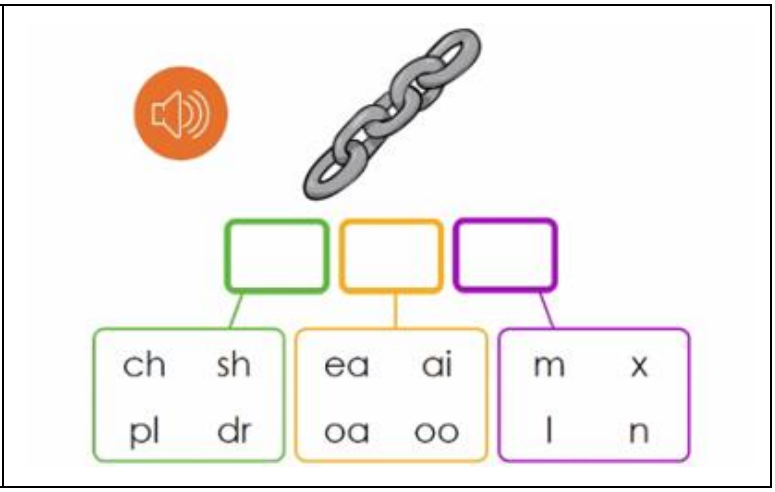
Decoding: Single Syllable	Decode the onscreen word and choose the picture that matches.	

Table 2.20. Specifications—Build Words: Single Syllable

Code	026
Specifications	Students build a given word using a set of letter options for each position in the word. The phonetically regular one-syllable word is given in audio and shown in a picture, and two or three empty boxes are shown into which students will pop a single letter or letter combination to spell the complete word. Because spelling is not the target of measurement, phonetically reasonable alternate spellings are not made feasible by the answer options. Consonant digraphs and blends are preserved intact. Where the medial vowel(s) can be separated from final consonant(s), there are three boxes with the second being for vowel letter(s). Where the vowel is inflected by final -l, -r, or -ng, or where a final silent *e* affects the vowel sound, the whole rime of the word is a single box. For each box, a set of four letters or letter combinations is included that are reasonable distractors (e.g., other vowels or vowel combinations; other initial consonant clusters; other whole rimes). Score is correct box completions per minute.
Item Pool	Up to 45 items presented in random order, each with two or three scorable boxes
Duration	2 minutes, speeded

CCSS Alignment	1.RF.3.b – Decode regularly spelled one-syllable words. 1.RF.3 and 2.RF.3 – Know and apply grade-level phonics and word analysis skills in decoding words.

Figure 2.16. Sample Item—Build Words: Single Syllable

Build Words: Single Syllable	Listen to a word and complete its spelling by choosing letters for each word component.	

2.5. Sentence Reading Fluency

When students can read a sentence silently with sufficient speed, accuracy, and literal comprehension, this indicates a level of proficiency with connected text beyond that indicated by isolated word reading. Several high-quality clinical assessments of reading include a measure in which students read isolated English sentences quickly and silently, then mark a quick semantic judgement. Examples include the Woodcock Johnson's Reading Fluency Task (Schrank et al., 2004) and the Test of Silent Reading Efficiency and Comprehension (Wagner et al., 2010). Such a measure draws from research indicating that stronger readers' comprehension is highly correlated to sentence-level silent reading fluency: when students do well on silent sentence reading, they are likely to read with good phrasing when reading aloud (Klauda & Guthrie, 2008). While word reading strongly predicts passage comprehension for weaker readers, silent sentence reading fluency has a tighter relationship to comprehension for stronger readers (Kim et al., 2011). In MAP Reading Fluency, the Sentence Reading Fluency measure is presented to all students to help discern readiness for oral passage reading.

Sentence Reading Fluency is a measure that students take at the beginning of the Adaptive Oral Reading test form. Their scores on this measure determine whether they route to Oral Reading Fluency or Foundational Skills. Other test forms are more constrained and dictate that all students assigned to that form will route to the same content track (e.g., everyone to Oral Reading Fluency or everyone to Foundational Skills). In that case, Sentence Reading Fluency may still be presented (along with Foundational Skills content), but it is not operating as the route determiner. It is just reported as a score, with NWEA guidance being that it is a great place to look to determine readiness for assigning passages.

Table 2.21. Specifications—Sentence Reading Fluency

Code	008
Specifications	Students read an onscreen sentence silently and choose the simple illustration that depicts its meaning from among four choices. Readability for single sentences cannot be scored by most readability formulae. Instead, educators with primary grade expertise reviewed sentences in item development to ensure that included words were either high frequency or decodable (phonetically regular) words. The target level of reading challenge is first grade with word count ranging from three to seven words. Score is correct selections over 2 minutes.
Item Pool	Up to 58 items presented in random order
Duration	2 minutes, speeded

Figure 2.17. Sample Item—Sentence Reading Fluency

Silent Sentence Reading	Choose the picture that matches the onscreen sentence.	She skates on ice. 

2.6. Language Comprehension

In the Simple View of Reading model (Gough & Tunmer, 1986), reading with comprehension is the product of decoding proficiency and language comprehension. Even if students' decoding skills are perfect, a weakness in understanding language—its vocabulary, structure, and syntax, as well as the ability to listen and make inferences based on what is heard—will suppress passage comprehension as students mature (Foorman et al., 2015; Lepola et al., 2016). While it is possible to assess passage comprehension directly once students can read connected text, it is critical to assess and build the language comprehension of students not yet reading independently. In MAP Reading Fluency, language comprehension is assessed without a decoding demand for two groups: (1) students on the Foundational Skills track (i.e., students not reading passages orally) and (2) students showing poor literal comprehension on lowest level passages (i.e., lowest Lexile levels).

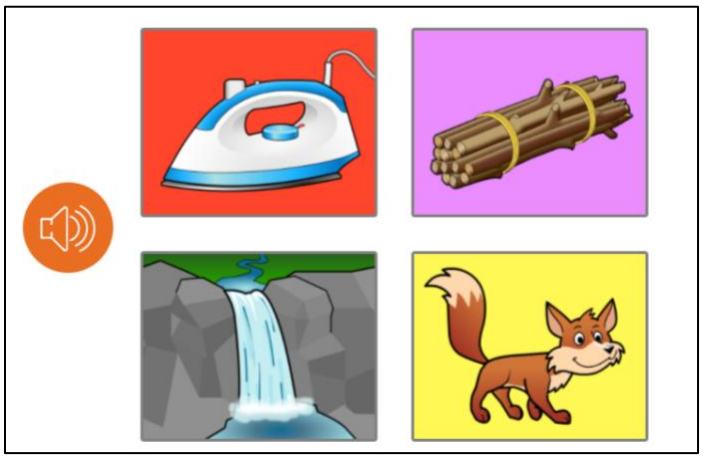
2.6.1. Picture Vocabulary

One aspect of a student's language comprehension is vocabulary knowledge. When a student selects a picture in response to a given vocabulary word, as in assessments such as the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007), receptive vocabulary is assessed. In MAP Reading Fluency, the focus is also on receptive or listening vocabulary, which is critically important for reading proficiency. When decoding an unfamiliar word, students who do not have the word in their listening vocabulary will not be able to determine if the decoded word makes sense in the context of the sentence or understand the author's intent (Biemiller, 2006). Research has shown that oral vocabulary from pre-kindergarten to first grade strongly predicts passage comprehension by fourth grade (Sénéchal et al., 2006; Scarborough, 1998; Cunningham & Stanovich, 1997).

Table 2.22. Specifications—Picture Vocabulary

Code	005
Specifications	Students choose the picture that matches the word given in audio only, with no onscreen text. Four pictures are presented onscreen. Vocabulary words are selected from a broad sample of curricular guides for kindergarten and first-grade vocabulary. Those not easily depicted in a simple illustration have been rejected. On a culled list, feedback was elicited in two cycles from educators with kindergarten and first-grade expertise and emergent bilingual expertise. Words with meanings that varied culturally or with confusing cognates in Spanish were removed. Numerically equal word lists were established for kindergarten and first grade separately, then combined. Score is the number of correct selections, with rate not being a factor.
Item Pool	15 items presented in randomly, from a pool of 42
Duration	Untimed

Figure 2.18. Sample Item—Picture Vocabulary



2.6.2. Listening Comprehension

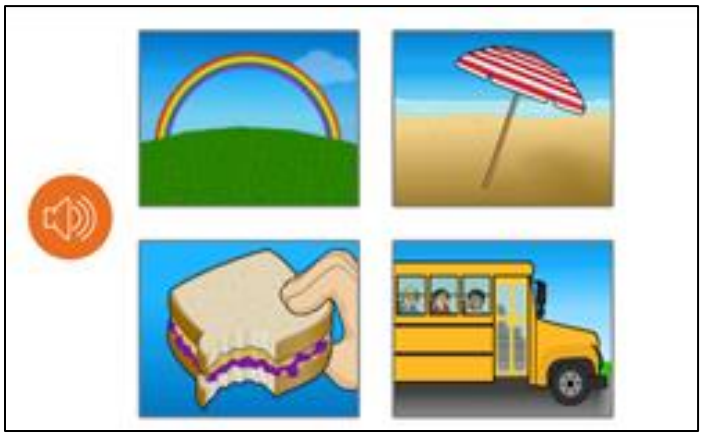
Language comprehension has been found to play a bigger role in later literacy achievement when it is measured using more complex measures that include grammar, the ability to define words, and listening comprehension than when measured using only simple vocabulary knowledge (Shanahan & Lonigan, 2010). MAP Reading Fluency includes both word and sentence-level language comprehension, in tandem.

Understanding the meaning of a sentence requires syntactic awareness. This involves understanding sentence structure (e.g., the use of grammatical rules) to ascertain meaning. Just as unfamiliar vocabulary will undermine fluent, automatic reading, so will unfamiliar syntactic structures in the text that students read. Researchers have found that syntactic awareness predicts passage comprehension (Catts et al., 2006; Mokhtari & Thompson, 2006; Nagy, 2007). Foorman et al. (2015) found that syntax, focusing on the sentence level, was a necessary component in a broader oral language factor that explained substantial variability in passage comprehension for grades K–2.

Table 2.23. Specifications—Listening Comprehension

Code	004
Specifications	Students choose the picture that matches the sentence given in audio only, without onscreen text. Four pictures are presented onscreen, with incorrect options including some semantic connection to the sentence (e.g., it includes one of the nouns in the picture) but that is clearly incorrect for a student comprehending the sentence. Audio playback is available. Two sets of sentences were developed, one for a kindergarten level and one for a first-grade level, then combined to form the measure. Each kindergarten sentence includes one or two grammatical constructions that can tax oral language comprehension in young students: prepositional and adverbial phrases, modifying clauses, verb modals, infinitives, and gerunds. In first-grade sentences, difficulty was increased by additional use of conceptual connectors (e.g., because, if), verbals and modals (gerunds, participles, should-could-would), more complex modifier structures (e.g., both direct and indirect objects; prepositional objects preceding verb), and more difficult vocabulary including homonyms requiring context. A significant constraint was that the sentence must be easily depicted by a simple illustration. Sentences failing this were thrown out. Feedback was elicited in two cycles from educators with kindergarten/first-grade expertise and emergent bilingual expertise.
Item Pool	15 items presented randomly from a pool of 37
Duration	Untimed

Figure 2.18. Sample Item—Listening Comprehension



2.7. Print Concepts

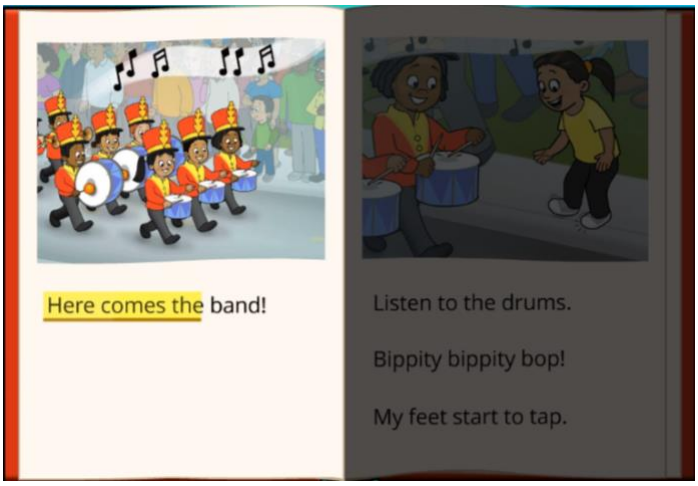
For the youngest beginning readers, an understanding of how print works may be important to gauge. Research indicates that students from lower socio-economic status enter school with weaker print awareness (Justice & Ezell, 2002). This matters: the National Early Literacy Panel found at least moderate correlations between knowledge about print conventions and concepts and later achievement in literacy (Shanahan & Lonigan, 2010). The CCSS frame these skills as “understanding of the organization and basic features of print” (Common Core State Standards Initiative, 2010, p. 15). Included are basic book skills like knowing where the cover is; concepts of word, including the understanding that print rather than pictures carry the language and how words are separated by spaces; and understanding of text directionality (e.g., page to page, left to right, top to bottom).

In MAP Reading Fluency, these print concepts are assessed within an interactive, multi-page electronic storybook format. A back and forth between student tasks and read-aloud by the narrator emulate the storybook context of traditional assessments of print concepts (e.g., Clay, 1989). Questions for the student are presented. After the student answers by touching part of the page (e.g., “Where should I start reading the words?”), the narrator reads the page aloud. A rolling highlight of the text being read reinforces the focus on print (Liao et al., 2020). Intervention research has shown that practices that increase attention to print can positively impact longer term literacy outcomes (Justice & Ezell, 2002, 2004; Piasta et al., 2012).

Table 2.24. Specifications—Print Concepts

Code	031-036
Specifications	Students choose the front cover of a book then answer questions about the inside text interspersed with a read aloud of the story text itself. Each storybook includes assessment of page-by-page reading, top-to-bottom reading, left-to-right directionality, return sweep across two lines of text, and differentiation of words by spaces. Each page includes both text and a picture. Responses are made by touching or clicking a location (e.g., the first word) on the two-page spread.
Item Pool	6 items within one storybook; 6 storybooks in pool
Duration	Untimed
CCSS Alignment	K.RF.1.a – Follow words from left to right, top to bottom, and page by page. K.RF.1.b – Recognize that spoken words are represented in written language by specific sequences of letters. K.RF.1.c – Understand that words are separated by spaces in print. K.RI.5 – Identify the front cover, back cover, and title page of a book.

Figure 2.19. Sample Item—Print Concepts

Print Concepts	Click within the open book to answer questions about print directionality and concept of word. Narrator reads the text aloud between tasks, using accompanying rolling highlight.	

2.8. Oral Reading

When students begin to read from connected text, fluency, or smooth and accurate reading, is introduced (National Reading Panel, 2000), which is a key focus for both instruction and assessment. Early focus on fluency sprung from the understanding that as students read words with more automaticity, they focus mental processing less on the decoding task and more on the task of producing meaning (LaBerge & Samuels,1974). Fluency in connected text pulls together the relationship described in the Simple View of Reading model (Gough & Tunmer, 1986) by enabling accurate word decoding to engage with language comprehension so that a student can integrate the two into meaning (Klauda & Guthrie, 2008).

Researchers have shown that WCPM scores can predict later success in reading, gauge students' response to instructional interventions, and indicate broader reading proficiency (Fuchs et al., 2001; Jenkins et al., 2007; Wayman et al., 2007). The tradition in curriculum-based measurement is to limit reading to one minute (Deno, 1985; Wayman et al., 2007). However, many have argued for assessments that include several key features from the more time-intensive approach of informal reading inventories. Such an approach includes reading whole short passages at varying levels, with word level accuracy explicitly scored instead of just rate (e.g., Leslie & Caldwell, 2006). The latter approach also allows for asking students comprehension questions after the passage, a design feature that many literacy scholars argue is essential to activating students' strongest reading behaviors (Samuels, 2007). Moreover, researchers have shown that supplementing reading rate scores with both accuracy and comprehension scores provides instructionally valuable diagnostic information and improved predictive validity (Valencia et al., 2010). In MAP Reading Fluency, all oral reading is scored for both rate and accuracy. For full oral reading passages, students are also scored on low-inference comprehension questions that follow the passage.

2.8.1. Oral Reading: Picture Book or Graphic Novel

All students taking the Adaptive Oral Reading form interact with a story in an onscreen "picture book" or "graphic novel" format, reading it aloud. For students in pre-kindergarten to second grade, a picture book format is used. Six pages are presented, with two side-by-side pages onscreen at one time. Each page has one or two sentences of text and a large picture supportive of meaning. This adopts the traditional book format used in research on shared book reading with younger children (e.g., Hargrave & Sénéchal, 2000; Mol et al., 2008).

For students in third grade and above, a "graphic novel" format is used. Three pages are presented, each with four cells of pictures supportive of the story's meaning. Words to read aloud are presented in a text box above the image within a cell. Use of this format for older students is designed to keep pace with the tremendous growth in the children's graphic novel market (Middaugh, 2019) and the growing research base on the increased engagement this format offers for older readers (Boerman-Cornell, 2016; Cornelius, 2020). In both formats, students choose when to use the button to proceed to the next page or indicate that they are finished with the last page. For students who cannot read connected text independently, audio captured might include decoding attempts at some words on the page or might include an invented "reading" of the pictures. No comprehension questions are associated with the picture book/graphic novel formats, and all are narrative stories.

Table 2.25. Specifications—Oral Reading: Picture Book

Code	013
Specifications	Each picture book was designed to be engaging for students across the primary grades and readable by beginning readers of connected text. They were developed to target low levels of text complexity, as measured by the Lexile Framework® for Reading, but also to provide significant picture support for students struggling to decode text independently. About 5–12 words appear on each page, along with a supportive illustration. Text and pictures were reviewed by experts in primary grades literacy assessment for quality and for age-appropriate content, form, and tone. Oral reading samples from the picture books are automatically scored for WCPM and accuracy. Human scoring for prosody is available via audio playback.

Table 2.26. Specifications—Oral Reading: Graphic Novel

Code	040
Specifications	Each graphic novel formatted story was designed to be engaging for students across the intermediate grades and readable by beginning readers of connected text. They were developed to target low levels of text complexity, as measured by the Lexile Framework® for Reading, but also to provide significant picture support for students struggling to decode text independently. About 5–15 words appear in each text box, appearing above a supportive illustration in the cell. Text and pictures were reviewed by experts in intermediate grades literacy assessment for quality and for age-appropriate content, form, and tone. Oral reading samples from the graphic novel formatted stories are automatically scored for WCPM and accuracy. Human scoring for prosody is available via audio playback.

Figure 2.20. Sample Item—Oral Reading: Picture Book

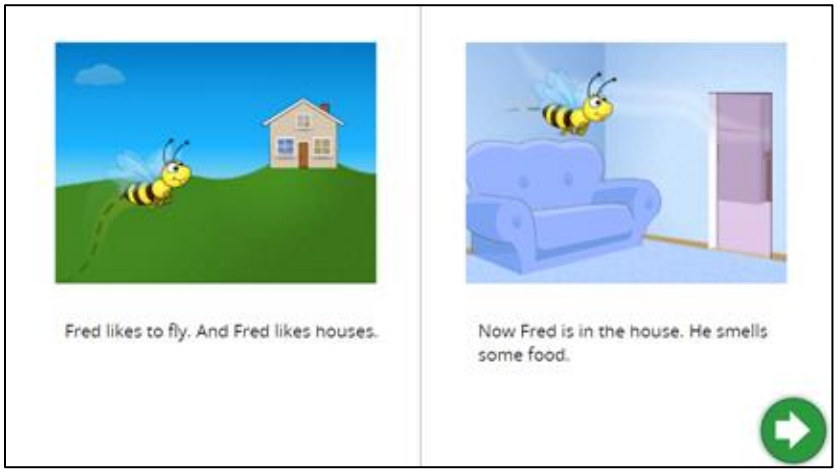


Figure 2.21. Sample Item—Oral Reading: Graphic Novel

Table 2.27 presents the traditional text Lexile Framework® for Reading readability measure and word count for each separate picture book or graphic novel format.[3] The Lexile Framework® for Reading provides a common scale for measuring text difficulty. A Lexile® measure is a number followed by an "L." The scale typically ranges from 0L to 1700L, although actual Lexile measures can be lower or higher. For example, a simple picture book might have a Lexile measure of 100L, while a college textbook might be measured at 1700L or higher (Lennon & Burdick, 2014). Lexile values below 0L are labeled as Beginning Reader (BR), which works like negative numbers (e.g., BR100L is higher than BR300L). The Lexile method for determining text complexity ratings includes four indicators, fed by quantitative metrics: structure, syntax, semantics, and decoding. Passages with the length and complexity necessary to support a comprehension quiz of six items were found to be infeasible to develop below 150L.

For picture book or graphic novel text, it is feasible to drop below 150L. In these formats, however, the Lexile measure is confounded by the pictures presented. Good illustrations play a role in supporting a student's experience of difficulty with all four Lexile factors but are not accounted for in the Lexile quantitative analysis. Because of this, the Lexile of picture book or graphic novel formats in MAP Reading Fluency was evaluated alongside qualitative evaluation of the degree of picture support to ensure that the experience would be appropriate for all levels of reader.

Table 2.27. Readability Measures and Word Count for Picture Book and Graphic Novel

Title	Lexile®	Word Count
Picture Book		
Bear on the Bus	120L	59
Jon Makes a Card	160L	61
Fred on a Hot Dog	160L	58
Walk Home with Best Friend	180L	47
Ken's Snow Day	190L	51
Duck in the Sink	210L	61
Jade's Grandma	230L	60
Star and Mom	310L	67

[3] This is different from the Lexile® oral reading measure reported on a MetaMetrics scale for English MAP Reading Fluency that accounts for student rate, student accuracy, and the text's oral readability.

Title	Lexile®	Word Count
Graphic Novel		
Planting Cereal	340	78
Rock Stars	280	70
Kickball Queen	400	76

2.8.2. Oral Reading: Passages and Comprehension Quiz

Students who have shown evidence of likely readiness for connected text reading are given passages, each with approximately 200 words, to read aloud followed by a series of six questions presented in a fixed order designed to require only literal or low-inference comprehension of the passage. Each set requires that no question is cued by a previous question, which necessitates a fixed order. For engagement, each set was also required to incorporate pictures into at least two questions, either as supplemental to the question stem or as answer options.

Table 2.28. Specifications—Oral Reading: Passages and Comprehension Quiz

Code	011, 014
Specifications	Students read the passage aloud and are alerted that questions about the passage will follow. The full text of the passage is presented onscreen, without the need for scrolling or page turning. Students use a button to indicate that they are finished. Each selected-response comprehension question appears and is read aloud by the narrator. Audio is available on answer options. Automatic scores for the oral reading include SWCPM and accuracy. Comprehension is reported as percent correct.

Figure 2.22. Sample Passage

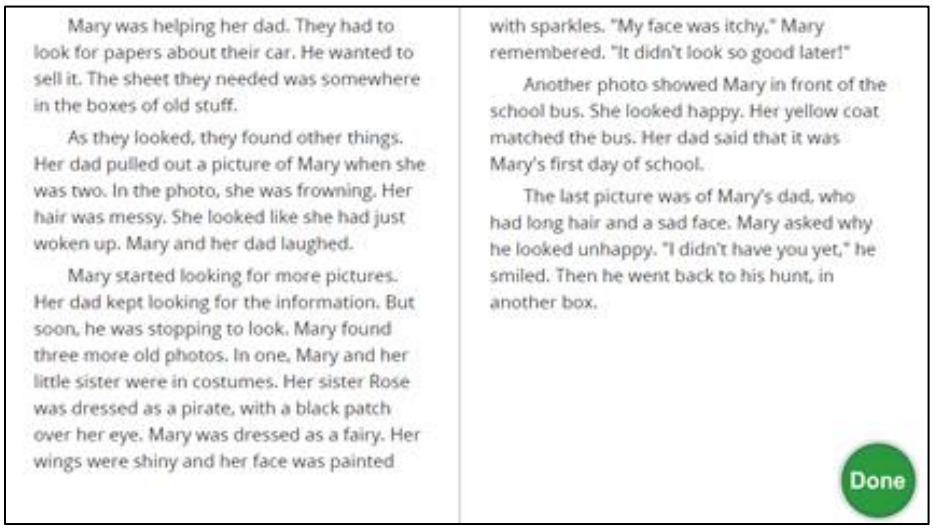
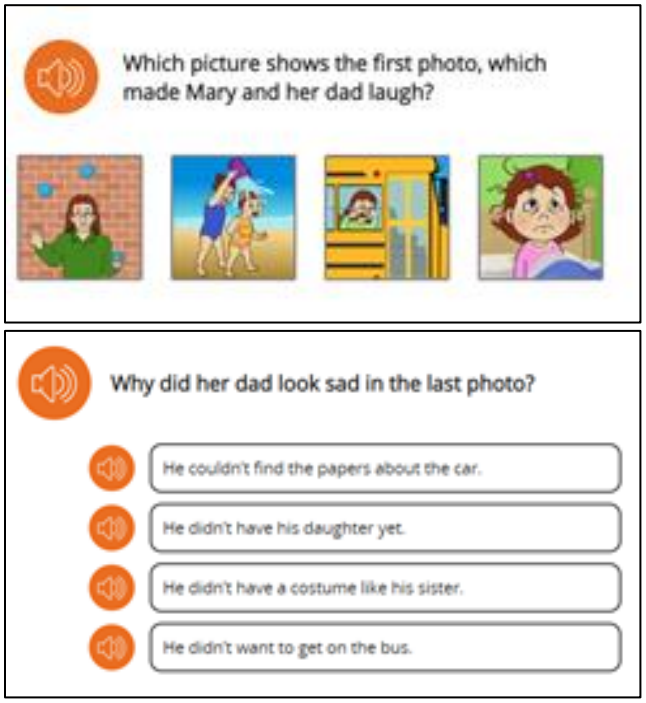


Figure 2.23. Sample Item—Oral Reading: Passage Comprehension Quiz



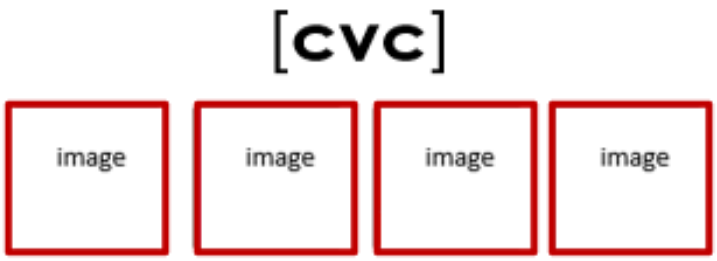
3. Content Development

For each component of foundational literacy included in MAP Reading Fluency, NWEA content specialists and external experts (i.e., professors and researchers with specialties in learning to read) reviewed the relevant research literature and the relevant academic standards and progressions, including the CCSS. After identifying critical domains and components, they determined the evidence necessary to demonstrate the knowledge and skills represented in each component. From these evidence requirements, development of a measure began with the design of an item template.

3.1. Item Template Creation and Review

For each measure within MAP Reading Fluency, ease of use by primary grade students made it imperative to design a set of items with maximum clarity and similarity of functioning. NWEA content specialists created item templates for each measure to ensure consistency across items in content scope, context, cognitive complexity, item format, graphics, and audio style. Figure 3.1 presents an example. An item format was designed and then populated repeatedly with content according to specifications to generate all items within a measure. Careful review of the item templates included determination of any corollary skills or understandings required to access the task. An iterative and collaborative design process was used by experts in early literacy to refine these templates, which were later used to design items across the scope of content defined by the measure.

Figure 3.1. Item Template Example—Decoding: CVC Words



Initial direction to student for this measure, in audio: *"Read the word, then click the matching picture."*

Basic item function: *Upon student clicking an image to respond, advance to next item.*

Inactivity rules*: After 10 seconds of inactivity, repeat initial direction. After student has been on the overall measure for 1 minute, advance to introduction of next measure.*

At the item template level, the approach and phrasing of the stem was determined and reviewed for best item construction practices (e.g., a full stem is not always repeated across sets of speeded measures). Stems were reviewed in two stages by experts in elementary grades literacy for adherence to best practices for young students. The following criteria were used.

Each stem should:

- Clearly connect a student to the concept, idea, or skill being assessed.
- Clarify the functionality of the task, where necessary.
- Use simple, age-appropriate vocabulary.
- Use simple syntax, including features such as present tense, active voice, and short sentence length.
- Be worded positively and directly.

The formal and structural approach of the answer options was also determined at the item template level. Determinations were set for whether answer options would be pictures, with or without audio; sentences or words with audio; or letters. Unless the inclusion of audio were to interfere with the evidence requirements, audio support would be included.

3.2. Item Writing and Review

Each item was written by NWEA content experts and multiple reviews, always within its set to maintain close match across items in functionality, clarity, and difficulty. Because stems were set at the template level, review at the item level focused on item assets (e.g., an audio and/or onscreen representation of a letter, sound, word, or sentence, possibly including a picture) and answer options (e.g., a letter, word, sentence, or picture, possibly with audio). The following criteria were used in the creation of the MAP Reading Fluency items.

Item assets should:

- Be engaging and relevant for Pre-K to Grade 5 students.
- Offer both visuals and audio, where feasible given evidence requirements.
- Be free of errors in grammar, usage, and mechanics.
- Be free of bias or sensitivity concerns.
- Be free of plagiarism or copyright infringement.

Answer options should:

- Have exactly one key.
- Represent typical student misconceptions where possible.
- Be feasible enough and close enough to require that students demonstrate the skill of interest in discerning the key.
- Compose a set that is not overlapping and does not include logical opposites, where possible, for sentences.
- Avoid null options such as “none of the above” or “all of the above.”
- Be visually clear and engaging, particularly for pictures.
- Be balanced in length, complexity, and grammatical structure for sentences and phrases.
- Use simple, age-appropriate vocabulary and syntax.
- Be engaging and relevant for Pre-K to Grade 5 students.
- Offer visuals and audio where feasible given evidence requirements.
- Be free of errors in grammar, usage, and mechanics.
- Be free of bias or sensitivity concerns.

3.3. Passage Development

Passages were developed at varying levels of text complexity, as gauged by the Lexile Framework® for Reading. Length could vary by grade level but was constrained by screen real estate; no passages requiring scrolling or page turning were included. Passages were reviewed in two stages by experts in primary grades literacy assessment for quality and age-appropriate language, content, form, and tone. They were reviewed separately for any issues with bias or sensitivity. In the first stage, passages were selected according to specific qualitative and quantitative criteria by NWEA content specialists:

- The passage is well written and engaging.
- The passage is age appropriate for students.
- The passage is free of bias, sensitivity, and fairness concerns.
- The passages focus on a variety of topics, including narrative and informational.
- The passage fits at the selected grade level when qualitative criteria are considered (e.g., levels of meaning or purpose; structure; language conventionality and clarity; knowledge demands).
- The passage fits onscreen without necessitating scrolling, with sufficient font size.
- The passage fits within a target Lexile measure.

In the second stage of passage review, NWEA publishing professionals reviewed passages for errors in grammar, usage, and mechanics; for issues of bias, sensitivity, and fairness; and to make sure the passages represent original material that does not infringe on any copyrights. Appendix A presents descriptive data for each passage used in MAP Reading Fluency.

3.4. Copyright and Permissions Review

The copyright and permissions specialist performs a review of all passages and items, ensuring that the item and asset content is free of plagiarism and that all trademark and Right of Publicity requirements are researched and documented. Phrases, strings of words, and images are searched online to ensure that items and item assets are free from plagiarism. Source materials provided by passage writers are also reviewed. When passages are factually based, writers must provide proof of their factual content. Writers attach documents and/or provide URLs showing where they obtained the information. The permissions team reviews these to make sure the sources have not been plagiarized.

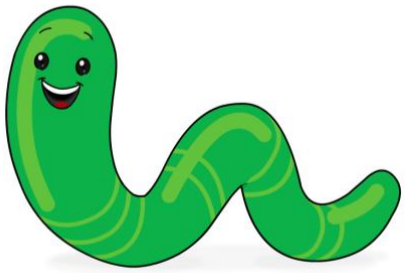
4. Test Administration and Security

MAP Reading Fluency is administered through the NWEA Comprehensive Assessment Platform. Access is housed on the same platform as MAP® Growth™, giving partners the convenience of a single login and common rostering. The actual MAP Reading Fluency application is delivered by LanguaMetrics™ through a separate application that students access through the student dashboard login. The MAP Reading Fluency benchmark and screening forms are best used seasonally in conjunction with MAP® Growth™ Reading.[4] Progress monitoring forms are designed for higher frequency administration.

4.1. User Experience

The assessment experience uses two avatar contexts, one for students in pre-K to Grade 2 and another for students in Grade 3 and above. Younger students are accompanied by a more effusive worm character (Wiggles), while older students are accompanied by a more age-appropriate bird (Swift the yellow warbler), as shown in Figure 4.1. Each is designed to be developmentally appropriate, although the item template per skill assessed remains stable within these varying contexts. Ultimately, the goal is to increase test engagement by giving students in Grade 3 and above a more mature avatar with age-appropriate language and pacing.

Figure 4.1. Avatars

Wiggles – Pre-K to Grade 2	Swift – Grades 3+
	

4.2. Practice Tests

Practice tests are available online year-round for students to familiarize themselves with the assessment. They provide the same access and functionality as the real MAP Reading Fluency tests. To take the practice tests, students log into their account to access the tests, entering their student username and password, and select between foundational skills and oral reading. Practice test specifics are as follows:

- Not adaptive
- Not scored
- No proctor control
- Available in any supported browser and any supported device
- Available for all supported grades
- Less than five minutes to complete the practice test

[4] Please refer to the *MAP Reading Fluency Administration Guidance Document* for more details on which test form to administer and how to leverage MAP Reading Fluency and MAP Growth Reading together to get a more complete picture of early literacy and development (NWEA, 2021b).

4.3. Administration Setup

To take the MAP Reading Fluency assessment, each student needs a computing device (PC/Mac/Chromebook/iPad) and an over-ear headset with a boom microphone. School staff should ensure that computers and headsets are operational and properly configured. Comprehensive and up-to-date guidance on technical setup can be found in the MAP Help Center, accessible from the top right of each page in the educator site.

Each PC, MAC, and Chromebook computer used for administration must have Google Chrome installed and be able to record audio from the test site. If this permission has not previously been granted for the device, an alert will prompt the user to do so. The most up-to-date version of Chrome is recommended, although earlier versions of the browser may be used if the minimum specifications are met. Full technical specifications outlining the minimum operating systems and browser versions are maintained by NWEA and available in the MAP Help Center. The MAP Reading Fluency iPad application is available free from the Apple Store. Students testing on an iPad log into the app using the credentials found on the educator site, just like students using the Chrome browser.

Prior to testing, students will have been enrolled and rostered into the MAP database and licensed to use MAP Reading Fluency. Students log in to a dedicated testing website or the iPad application using a username and password that can be assigned by the school or generated by the MAP system.

All administration instructions are presented by audio within the test. A microphone check ensures that the recording equipment is functioning at the time of the test. It is essential that students use an external microphone for oral reading measures, and all tests require audio output. To test audio input and output levels, school staff may log into the educator site or student site and use the Check Equipment module to record and playback test audio. Prior to testing, it is recommended that each device be checked through this module to minimize the likelihood of having to adjust settings when students are waiting to take the test. When testing on an iPad, the audio check is found on the login page before logging in.

4.4. Managing Students and Test Sessions

Because all content presentation, response capture, and scoring are done automatically by the system, MAP Reading Fluency can be administered in a group setting. A single adult proctor can oversee a classroom full of students simultaneously taking MAP Reading Fluency. However, smaller groups with 8–10 students are recommended to improve background noise conditions and promote easier classroom management. Students should be spread out as much as is practical. High background noise can lead to audio records that the speech engine cannot score. If the group size is greater than 10 students, it is recommended to have two adults present. This allows one adult to assist an individual student in case of technical or personal difficulty while the other oversees the class.

An optional mouse screening activity can be administered prior to a student test session. This activity challenges students to respond in a manner similar to the test and ensures that they can operate the equipment and respond appropriately to the instructions and prompts. The mouse skills check is recommended once at the beginning of pre-kindergarten, kindergarten, or first grade, unless the student is testing on an iPad.

4.5. Pausing, Resuming, and Discarding In-progress Tests

Students typically take 20–30 minutes to complete the MAP Reading Fluency assessment. Completion within one sitting is recommended but not required. If a student needs to take a break during the test, three mechanisms support this:

1. A pause button that appears during instruction screens.
2. A user-initiated “start recording” button that appears before each oral reading attempt, which may be left unclicked during a brief break. This button is a large green circle in the middle of the screen.
3. Closing the browser window, which will automatically pause the test and allow it to be resumed later by logging back in.

Any in-progress test session that has been paused, actively or by default (e.g., power failure), will resume automatically when the student logs back in. At the discretion of the teacher, an in-progress test can be discarded, and the student will be allowed to start the test from the beginning. A teacher makes this selection from the Proctor Dashboard or Assignments page. While students can complete equivalent test forms up to three times, the system only maintains one active session at a time. This session should be discarded if the teacher wants the student to begin again. Based on the content presentation logic, students will likely see some of the same content on a second attempt.

4.6. Test Security

Inadequate security procedures pose a risk to assessment systems. Violations of test security may compromise the integrity of results and call into question the trustworthiness of information. A common criticism of test security relative to adaptive tests is that some tests do not use sufficiently large item pools to ensure that content on the test cannot be “poached” by groups of students or educators who memorize, compile, and share large numbers of items. However, well-designed, adaptive tests such as MAP Reading Fluency that draw from large item pools offer several advantages for ensuring test and item security. The MAP Reading Fluency systems leverage the following security advantages:

- Items are only available to authorized users of the system.
- Passages will not be repeated until all the passages in a specific item pool are exposed to the user.
- Items are randomized and presented to users.
- Every student must log in to their individual account and can start a test assigned by the system or by the educator.
- Item types are not stored/cached locally. Responses are stored in secure servers before presenting the next item type to the student.

The processes and tools provided in Table 4.1 are also used to ensure that the integrity of the tests are not jeopardized, providing educators and students a positive and reliable user experience.

Table 4.1. Test Security Before and During Testing

Before test administration	• Rostering of student and educator data through secure system applications. • Only specific user roles, approved and authorized within the district and school, can log into the system to access test administration features.
During test administration	• Students can only access the test assigned by the system or educator. • Students can only have one active session and will be logged out if they try to open another session.

4.6.1. Assessment Security

All transmissions of testing and response data are encrypted and secured using TLS 1.2 AES 256 encryption methods. Test data are stored in highly secure Amazon data centers located in the continental U.S. operating with redundant power, internet, and backup systems powered by diesel generators. All servers, disk storage, and network infrastructure are redundant, protecting against unavailability due to a single hardware failure. NWEA operates in multi-availability zones by Amazon with data replication for failover if one data center becomes inoperable. Personally identifiable student information is encrypted at rest in the systems. More information can be found at https://legal.nwea.org/map-growth-information-security-whitepaper.html.

4.6.2. Role-Based Access

Access management is a critical function for maintaining test security. MAP Reading Fluency uses role-based access security controls that allow partners to segregate duties in their MAP Reading Fluency accounts and grant only the amount of access to users needed to perform their jobs. This allows partners to control what actions and data individuals have access to. When planning partners' access control strategy, MAP Reading Fluency supports granting users the least privilege to perform their work. Each role in MAP Reading Fluency has specific permissions that control levels of access to implementation, configuration, data management, testing, and reporting tasks. Each user has a unique username to which one or multiple roles can be assigned. Only certain roles can create or modify student profiles, which limits the ability to change student information. More information can be found at https://teach.mapnwea.org/impl/QRM2_Roles_and_Responsibilities_QuickRef.pdf.

5. Scoring and Reporting

All student responses are scored automatically by the MAP Reading Fluency software. The reported outcomes of each measure are presented in Table 5.1. The Oral Reading measures that yield SWCPM scores scored by the LanguaMetrics software embedded in the test engine. All other measures are selected-response and are scored dichotomously, either correct or incorrect, at the item level by the test engine. Raw scores and number of items attempted are reported in the reporting site. A performance level is also assigned in each domain: *Exceeds Expectation*, *Meets Expectation*, *Approaching Expectation*, and *Below Expectation*.

Students can obtain Foundational Skills scores in one of two ways: (1) they are routed to the Foundational Skills track if they are not yet ready to independently read passages aloud, or (2) their teacher assigns them to take a Foundational Skills form. In contrast, students can obtain Oral Reading Fluency scores in one of two ways: (1) they are routed to the Oral Reading Fluency track if they pass the Sentence Reading Fluency measure and progress to independent passage reading, or (2) their teacher assigns them to take the Passages Only form.

Table 5.1. Scoring Method and Reported Outcomes by Measure

Domain	Measure	Code	Scoring Method	Reported Outcomes
Phonological Awareness	Rhyme Completion	030	Dichotomously scored at the item level	Number correct, number attempted, and scaled domain score
	Counting Syllables	017		
	Onset -Rime Blending	018		
	Initial Sound Matching	001		
	Blending Phonemes	019		
	Phoneme Counting	020		
	Phoneme Addition/Deletion	021		
	Phoneme Substitution	022		
Phonics & Word Recognition	Letter Knowledge	002	Dichotomously scored at the item level	Number correct, number attempted, and scaled domain score
	Letter-Sound Fluency	003		
	Build Words: One Letter	024		
	Word Families: Initial Letter	023		
	Decoding: CVC	007		
	Build Words: CVC	025		
	Decoding: Single Syllable	027		
	Build Words: Single Syllable	026		
	Sentence Reading Fluency*	008		
Language Comprehension	Picture Vocabulary	005	Dichotomously scored at the item level	Number correct, number attempted (typically all 15 are attempted), and scaled domain score
	Listening Comprehension	004		
Print Concepts	Print Concepts	031–036	Dichotomously scored at the item level	Number correct and number attempted (typically all 6 are attempted)
Oral Reading	Oral Reading: Picture Book/Graphic Novel	013/040	LanguaMetrics speech scoring software	SWCPM; percent accuracy
	Oral Reading: Passages	011		
	Oral Reading: Passage Comprehension Quiz	014	Dichotomously scored at the item level	Percent correct out of 6 for each quiz

*Even though Sentence Reading Fluency is a Phonics & Word Recognition measure, it does not contribute to the Phonics & Word Recognition domain score.

5.1. Foundational Skills

Foundational Skills includes measures in the Phonological Awareness, Phonics & Word Recognition, and Language Comprehension domains. Phonological Awareness and Phonics & Word Recognition are assessed with a series of discrete, timed measures focusing on a single skill. Zone of proximal development (ZPD) levels are achievable from a series of related measures administered from each domain progression, as shown in Table 5.2. Students move through each progression based on their demonstrated ability, receiving 3–6 measures based on adaptive branching criteria in the test.

A ZPD level and accompanying performance level are achieved, as outlined in Table 5.3. Performance levels are color-coded as blue, green, yellow, or red (i.e., *Exceeds Expectation*: blue, *Meets Expectation*: green, *Approaching Expectation*: yellow, and *Below Expectation*: red). Performance levels are assigned at the domain level (i.e., at the level of the entire progression) by comparing the observed ZPD to grade-level expectations. Grade-level expectation is set at Level 1 in fall for kindergarten and Level 4 in winter for Grade 1. Beginning in the spring of Grade 1, the grade-level expectation is that students have moved out of the Foundational Skills track and into Oral Reading Fluency.

Table 5.2. ZPD Levels for Phonological Awareness and Phonics & Word Recognition

Phonological Awareness					
Level 0: Rhymes and Syllables (Introduce)	**Level 1: Rhymes and Syllables**	**Level 2: Initial Sounds**	**Level 3: Blending Phonemes and Segmenting**	**Level 4: Phoneme Manipulation**	**Level 5: Phoneme Manipulation (Reinforce)**
Rhyme Completion *Measures phonological rhyme identification skills*		**Onset-Rime Blending** *Measures initial phoneme blending skills*	**Blending Phonemes** *Measures phoneme blending skills*	**Phoneme Addition/Deletion** *Measures phoneme manipulation skills*	
Counting Syllables *Measures phonological syllable segmenting skills*		**Initial Sound Matching** *Measures initial phoneme identification skills*	**Phoneme Counting** *Measures phoneme segmenting skills*	**Phoneme Substitution** *Measures phoneme manipulation skills*	
Phonics & Word Recognition					
Level 0: Letters and Sounds (Introduce)	**Level 1: Letters and Sounds**	**Level 2: Letters in Words**	**Level 3: CVC Words**	**Level 4: One-Syllable words**	**Level 5: One-Syllable words (Reinforce)**
Letter Knowledge *Measures letter identification knowledge*		**Build Words: One Letter** *Measures letter sound decoding skills in word*	**Decoding: CVC** *Measures early word decoding skills*	**Decoding: Single Syllable** *Measures word decoding skills*	
Letter-Sound Fluency *Measures letter sound correspondence knowledge*		**Word Families: Initial Letter** *Measures letter sound decoding skills in words*	**Build Words: CVC** *Measures early word encoding skills*	**Build Words: Single Syllable** *Measures word encoding skills*	

Table 5.3. Performance Expectations by ZPD Level

Administration	ZPD Level					
	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
Kindergarten						
Fall	*Approaching*	*Meets*	*Exceeds*			
Winter	*Below*	*Approaching*	*Meets*	*Exceeds*		
Spring	*Below*		*Approaching*	*Meets*	*Exceeds*	
Grade 1						
Fall	*Below*		*Approaching*	*Meets*	*Exceeds*	
Winter	*Below*			*Approaching*	*Meets*	
Spring	*Below*				*Approaching*	
Grade 2						
Fall	*Below*				*Approaching*	
Winter	*Below*					
Spring	*Below*					
Grade 3						
Fall	*Below*					
Winter	*Below*					
Spring	*Below*					

The Language Comprehension domain includes the Picture Vocabulary and Listening Comprehension measures. It is assessed within the Foundational Skills section of the test and for students who proceed to passages but struggle to understand passages at the lowest Lexile levels. Each measure presents 15 items to the student, drawn randomly from a larger pool. Performance on each measure is assigned a performance level based on the number correct out of 15, as shown in Table 5.4.

Table 5.4. Performance Expectations for Language Comprehension

Grade	Number Correct of 15			
	Below Expectation	***Approaching Expectation***	***Meets Expectation***	***Exceed Expectation***
K	6 or less	7--8	9–11	12+
1	8 or less	9–11	12+	–
2	8 or less	9–11	12+	–
3	8 or less	9–11	12+	–

5.2. Oral Reading Fluency

A summary of student performance across all oral passage reading attempts is provided across three instructionally important dimensions of oral reading: oral reading rate (i.e., SWCPM), decoding accuracy, and passage comprehension. Valencia et al. (2010) have shown that providing data on each of these components offers greater predictive validity than use of SWCPM alone. Moreover, the student profiles of at-risk readers vary across these dimensions in ways that make a one-size-fits-all instructional approach ineffective: some students struggle with accuracy only, while others have a high rate of accuracy but low comprehension. Each profile calls for a different set of instructional emphases (Valencia & Buly, 2004). For each permutation of strengths and difficulties, MAP Reading Fluency refers teachers to an individually assigned recommendation for instructional focus and strategies.

5.2.1. Oral Reading Rate

Oral reading rate, using the metric of SWCPM, is considered based on the expectation levels in Table 5.5. Specifically, based on published norms for WCPM scores (Hasbrouck & Tindal, 2017), Table 5.5 presents the minimum thresholds (i.e., minimum WCPM) for reaching the *Meets Expectation* performance level relative to grade-level text. Table 5.6 presents the ranges for all performance levels. Students meet expectation if their overall SWCPM exceeds the minimum WCPM for a given grade and term. If students struggle to understand a grade-level passage, they will get an easier (lower Lexile) passage. If their fluency level on the easier passage surpasses a performance level boundary by 10 WCPM, the higher performance level will be achieved. Increased instructional intensity is suggested for students reading at a rate significantly below expected levels.

Table 5.5. Minimum Thresholds for *Meets Expectation*

	Minimum WCPM for *Meets Expectation**		
Grade	**Fall**	**Winter**	**Spring**
K	N/A		
1	N/A	29	60
2	50	84	100
3	83	97	112
4	94	120	133
5+	121	133	146

*N/A = not applicable; no oral reading expected.

Table 5.6. Performance Levels by SWCPM Ranges

		SWCPM Ranges		
Grade	**Performance Level**	**Fall**	**Winter**	**Spring**
K	*Exceeds Expectation*	Any oral reading		
	Meets Expectation	0 (no expectation)	0 (no expectation)	0 (no expectation)
	Approaching Expectation	–	–	–
	Below Expectation	–	–	–
1	*Exceeds Expectation*	9+	59+	91+
	Meets Expectation	0 (no expectation)	29–58	60–90
	Approaching Expectation	–	16–28	34–59
	Below Expectation	–	0–15	0–33
2	*Exceeds Expectation*	84+	109+	124+
	Meets Expectation	50–83	84–108	100–123
	Approaching Expectation	36–49	59–83	72–99
	Below Expectation	0–35	0–58	0–71
3	*Exceeds Expectation*	104+	137+	139+
	Meets Expectation	83–103	97–136	112–138
	Approaching Expectation	59–82	79–96	91–111
	Below Expectation	0–58	0–78	0–90
4	*Exceeds Expectation*	125+	143+	160+
	Meets Expectation	94–124	120–142	133–159
	Approaching Expectation	75–93	95–119	105–132
	Below Expectation	0–74	0–94	0–104

Grade	Performance Level	SWCPM Ranges		
		Fall	Winter	Spring
5+	*Exceeds Expectation**	–	–-	–
	Meets Expectation	121+	133+	146+
	Approaching Expectation	87–120	109–132	119–145
	Below Expectation	0–86	0–108	0–118

**Exceeds* is not reported if above-grade level passages are not provided.

5.2.2. Decoding Accuracy

Across all passages, a threshold of 95% is used to highlight students whose decoding accuracy may be limiting fluency and understanding. In a comprehensive review of how reading accuracy interacts with instructional text leveling, Allington et al. (2015) find that a minimum of 95% accuracy predicted significant increases in both engagement and comprehension. Specifically, Table 5.7 presents the boundaries for performance levels for decoding accuracy, which is classified according to ranges of percent accuracy on grade-level text. *Exceeds Expectation* is only achievable on grade-level text or higher. For below-grade-level text, *Exceeds Expectation* is replaced with *Meets Expectation* for students achieving 98% accuracy or higher.

Table 5.7. Performance Levels for Decoding Accuracy Based on Percent Accuracy

Performance Level	Decoding Accuracy
Exceeds Expectation	98–100%
Meets Expectation	95–97%
Approaching Expectation	90–94%
Below Expectation	0–89%

5.2.3. Passage Comprehension

Across all passages, answering five of the six passage comprehension quiz items correctly is used as a threshold for demonstrating basic understanding of the passage. Passage comprehension performance levels are assigned based on the most difficult text for which a student demonstrated understanding by answering at least five of six items correctly. Above-grade text produces *Exceeds Expectation* designations, and below-grade is *Approaching* or *Below* depending on the discrepancy from the grade level.

5.2.4. Lexile® Oral Reading Measure

MAP Reading Fluency also reports a Lexile Framework® for Oral Reading score (MetaMetrics, 2021). The student Lexile oral reading measure is generated using a combination of three factors: the student's oral reading rate, the student's oral reading accuracy, and the text's oral readability. The student score is presented in the context of typical oral readability for grade-level texts to allow comparing of student oral reading proficiency to grade-level demands.

In the Common Core era, elementary students are often asked to read in increasingly complex texts, including challenging grade-level texts, regardless of a "best match" level. As Shanahan et al. (2016) note, readers "build muscle" in reading by working with more challenging texts. Given this context, the possible gap between the Lexile oral readability of typical grade-level text and the student Lexile oral reading measure indicates the degree of instructional support required to help students work with grade-level text. Research indicates that where significant support is designed into instruction, all students can benefit from experiences with texts that might otherwise be characterized as "too hard" (Stahl & Heubach, 2005; Allington et al., 2015).

5.3. Individual Student Reports

The Individual Student Report shows all scores achieved on a given assessment, including profile statements that are linked to suggested instructional next steps. Each completed test can be reviewed by choosing the test date from the dropdown on the individual's page, which is accessed by selecting a student from the class list on the Student Matrix. All other MAP Reading Fluency reports are based on the data in the Individual Student Report.

Figure 5.1 presents the report layout for a student who has read passages aloud and answered comprehension questions, and Figure 5.2 presents the layout for students who have taken foundational skills measures within the Adaptive Oral Reading test format. Both examples show data of a student who has been flagged. Students who read passages receive a summary of their performance across the three sub-scores of oral reading rate, decoding accuracy, and passage comprehension, along with links to instructional recommendations. For students with Foundational Skills results who did not attempt oral reading, an analogous summary of student performance and instructional readiness is provided with linked suggestions for instructional focus based on the observed ZPD and oral language levels.

Figure 5.1. Sample Individual Student Report—Oral Reading

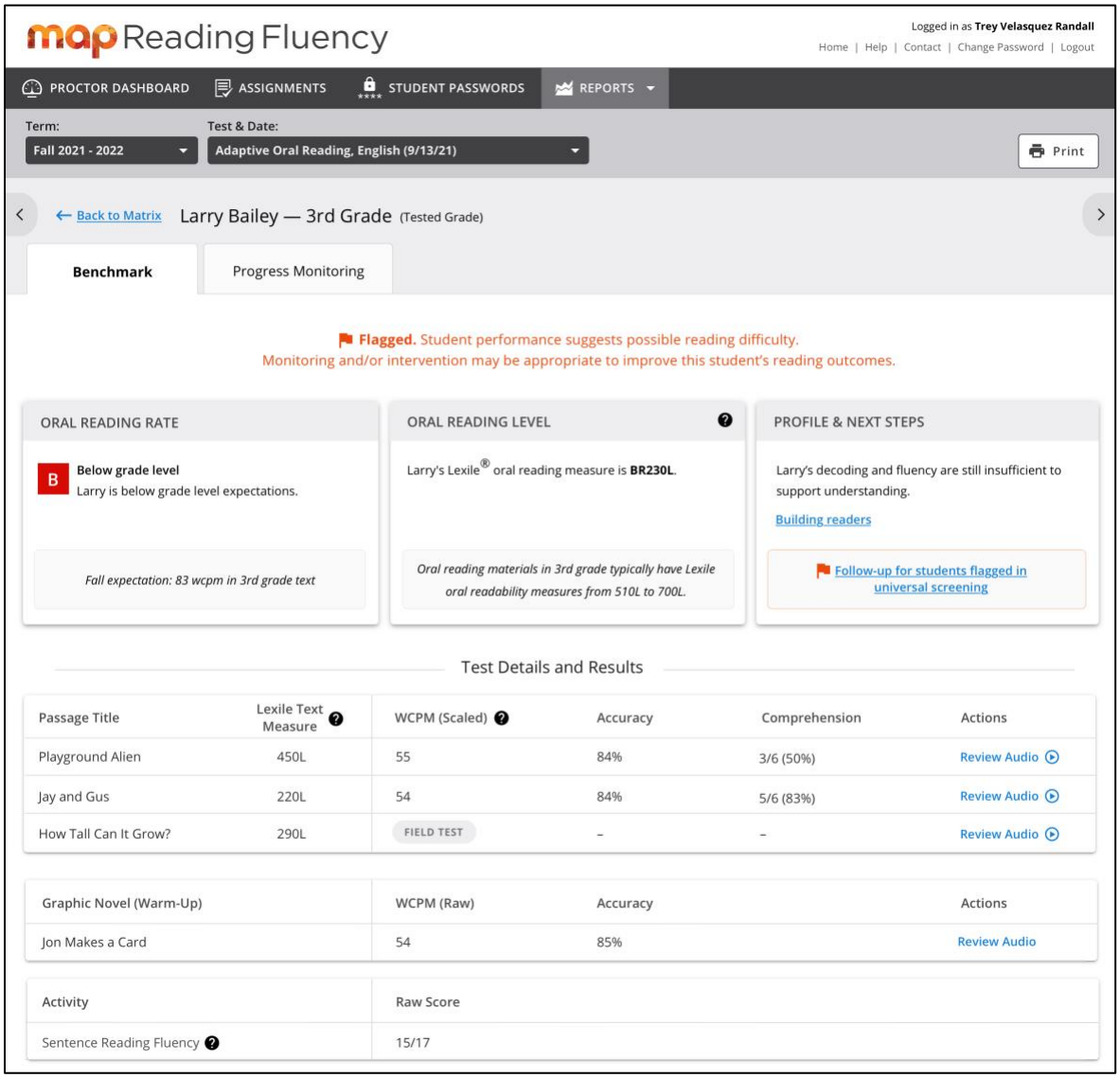
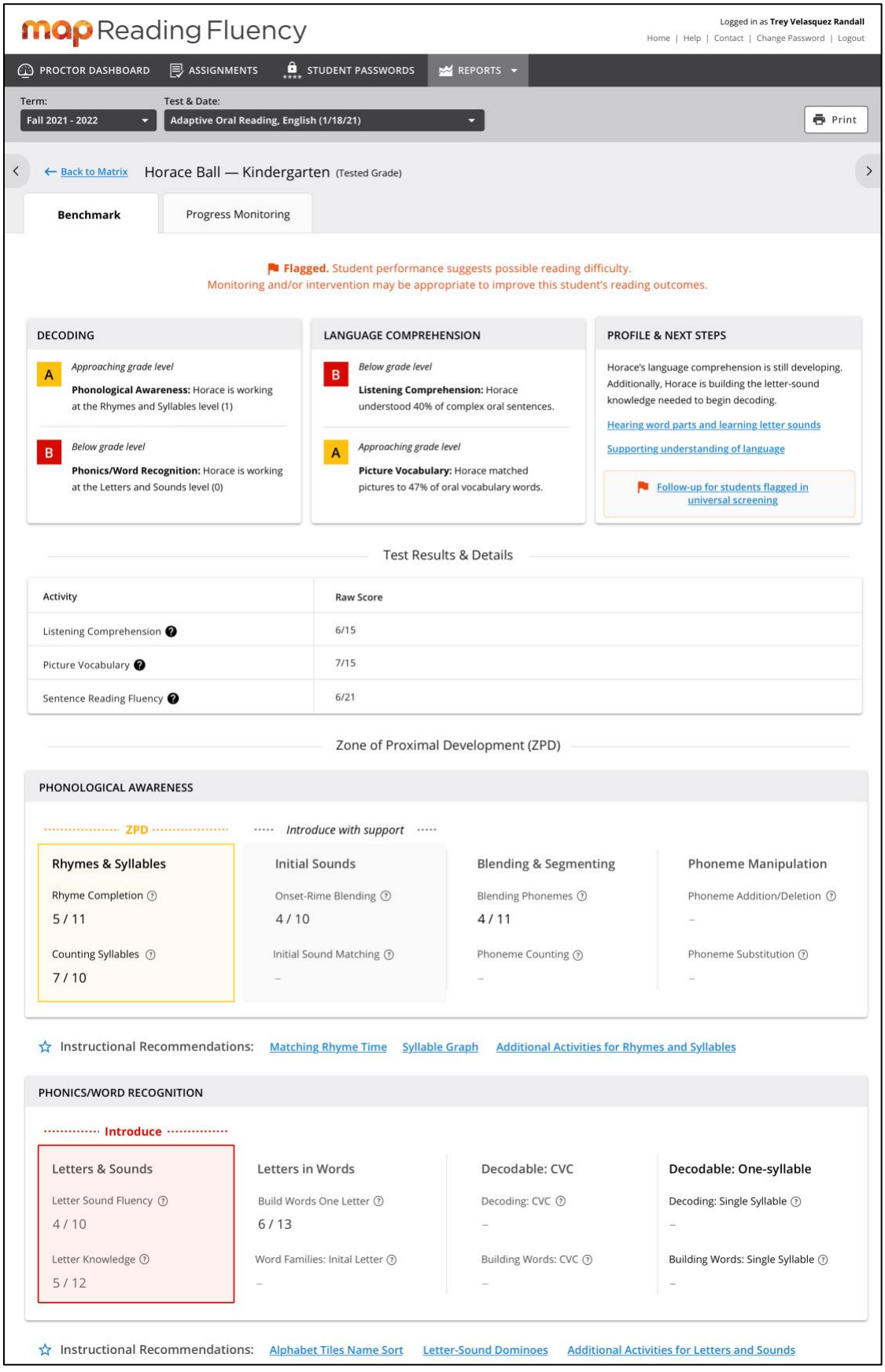


Figure 5.2. Sample Individual Student Report—Foundational Skills



5.4. LanguaMetrics' Speech Scoring Technology

NWEA partnered with LanguaMetrics to develop the speech scoring engine that scores the Oral Reading measures with WCPM reported outcomes (i.e., Oral Reading: Picture Book and Oral Reading: Passages). LanguaMetrics' team of scientists and engineers has pioneered the application of speech scoring science to education technology. The speech scoring technology has complex components such as acoustical models and speech recognizers. Acoustical models combine with a data dictionary and the speech recognizer to score speech. Acoustical models are based on thousands of speech samples that are run through modeling tools and optimization tools to produce the resulting model. The model is a statistical representation of all the details of speech associated with the population of the samples used. The broader the population, the less accurate the model. Therefore, the population was defined as narrowly as possible to produce the most accurate acoustical model possible. This is a key factor in the accuracy of the MAP Reading Fluency scoring mechanism because it uses an acoustical model created specifically for young student's voices.

The science within these components relies on a concept from Bayesian statistics known as hidden Markov models (HMMs) that are used in speech science to better understand the audio signal being recognized and scored. Every language has observable and discrete patterns based on the rules of the language. With HMMs, these rules are leveraged to lower the possibility of errors in recognition. For example, in the English language, the probability of the letter B coming after T is extremely low. Therefore, when speech is being recognized, the speech recognizer paired with the acoustical model is better equipped to return results that make sense for the targeted language. Additional data elements are returned by the speech recognizer, including confidence levels for both sentences and words, and various phoneme-level scores. These data are analyzed to create algorithms at the application level that are used to evaluate the reading of connected text.

Measuring and scoring the speech of young readers is far more challenging than typical speech recognition applications and requires the software to be able to accommodate young readers' wide degree of decoding skills and oral reading fluency development. Therefore, many of the words that need to be scored are not at the same level of articulation quality that speech recognizers would normally require to score with sufficient accuracy. Young readers may also skip words, repeat words, skip sentences, pause or remain silent for periods, and restart themselves at seemingly random positions in the text.

MAP Reading Fluency algorithms leverage statistical output from the reading grammar and the speech recognizer. These algorithms form the basis for the WCPM calculation and require calibration to achieve the desired precision and accuracy. The desired level of precision and accuracy is that the software return an oral reading evaluation that is in line with that of a typical teacher. (See Section 6.3.4. for results from human-machine agreement studies.)

6. Technical Characteristics

6.1. Student Sample

Data for most of the technical report analyses were collected during the 2020–2021 school year.[5] Table 6.1 presents the number of students who took the MAP Reading Fluency assessment in 2020–2021 by grade, term, and track (i.e., Foundational Skills or Oral Reading Fluency). From Fall 2020 through Spring 2021, 407,964 students received Foundational Skills scores and 277,920 students received Oral Reading Fluency scores. The All Students total is less than the sum of Foundational Skills and Oral Reading Fluency totals because students could have both a Foundational Skills score and Oral Reading Fluency score in one term. Students could also take different tracks throughout the year. For example, the same student might have had a Foundational Skills score in the fall but an Oral Reading Fluency score in the winter and spring.

Table 6.2 presents the demographic composition of the MAP Reading Fluency user population during the 2020–2021 school year and the U.S. pre-K–12 public school population based on data from 2020–2021 (U.S. Department of Education, n.d.). It also presents the number of MAP Reading Fluency students in each geographic division based on the Census Bureau (U.S. Census Bureau, 2016). The MAP Reading Fluency *N*-counts differ from the totals in Table 6.1 because Table 6.2 includes unique students only once across terms. Based on Table 6.2, the demographic composition of the MAP Reading Fluency test user population is reasonably close to that of the U.S. pre-K–12 public school population. The table also shows that the MAP Reading Fluency test user population comes from all nine of the Census Bureau's geographic divisions.

Table 6.1. Number of Students by Grade and Term

Grade	*N* Students who took MAP Reading Fluency		
	Fall 2020	Winter 2021	Spring 2021
All Students			
K	76,445	71,921	98,746
1	97,614	90,267	102,795
2	87,525	72,332	90,579
3	40,410	35,490	36,692
4	18,089	15,704	16,196
5	12,802	10,909	10,534
6	1,092	1,312	1,172
7	637	257	207
8	238	174	159
9	1	1	0
10	1	2	3
Unknown/Missing	79,931	79,005	81,066
Total	414,785	377,374	438,149

[5] The datasets for Table 6.1 and Table 6.2 were further cleaned based on the applicable data inclusion rules for each analysis in this technical report.

Grade	*N* Students who took MAP Reading Fluency		
	Fall 2020	Winter 2021	Spring 2021
Foundational Skills			
K	75,605	69,972	95,045
1	91,060	73,286	77,428
2	60,006	33,898	31,958
3	12,765	8,784	7,370
4	798	475	400
5	321	293	203
8	1	1	0
Unknown/Missing	51,769	47,407	45,198
Total	292,325	234,116	257,602
Oral Reading Fluency			
K	75,605	69,972	95,045
1	91,060	73,286	77,428
2	60,006	33,898	31,958
3	28,974	27,624	30,169
4	17,732	15,481	15,995
5	12,613	10,766	10,443
6	1,092	1,312	1,172
7	637	257	207
8	238	174	159
9	1	1	0
10	1	2	3
Unknown/Missing	29,555	33,600	38,464
Total	134,320	156,859	203,458

Table 6.2. Demographic Characteristics

Demographic Subgroup	MAP Reading Fluency Student Population, 2020–2021						U.S. Pre-K–12 Public School Population, 2020–2021	
	All Students		Foundational Skills		Oral Reading Fluency			
	N	%	*N*	%	*N*	%	*N*	%
Total	568,758	100.0	407,964	100.0	277,920	100.0	49,356,945	100.0
Sex								
Female	229,041	40.3	165,271	40.5	111,657	40.2	23,986,757	48.6
Male	238,844	42.0	175,036	42.9	113,663	40.9	25,302,030	51.3
Unknown/Missing	100,873	17.7	67,657	16.6	52,600	18.9	68,158	0.1
Race/Ethnicity								
American Indian/Alaska Native	5,163	0.9	4,043	1.0	1,971	0.7	459,906	0.9
Asian/Pacific Islander	19,852	3.5	14,476	3.5	11,315	4.1	2,675,163	5.4
Black/African American	91,279	16.0	72,087	17.7	40,001	14.4	7,402,625	15.0
Hispanic/Latino	116,193	20.4	81,848	20.1	56,567	20.4	13,808,085	28.0
Multi-Ethnic/Two or More Races	21,039	3.7	15,652	3.8	9,846	3.5	2,212,631	4.5
Native Hawaiian/Other Pacific Islander	765	0.1	494	0.1	402	0.1	180,660	0.4
White	185,375	32.6	132,901	32.6	90,978	32.7	22,549,717	45.7
Unknown, Other, or Missing	129,092	22.7	86,463	21.2	66,840	24.1	68,158	0.1

Demographic Subgroup	MAP Reading Fluency Student Population, 2020–2021						U.S. Pre-K–12 Public School Population, 2020–2021	
	All Students		Foundational Skills		Oral Reading Fluency			
	N	%	*N*	%	*N*	%	*N*	%
Census Division								
New England	17,136	3.0	12,649	3.1	7,974	2.9	–	–
Middle Atlantic	17,810	3.1	11,619	2.8	8,663	3.1	–	–
East North Central	73,563	12.9	52,628	12.9	37,141	13.4	–	–
West North Central	23,932	4.2	17,498	4.3	10,828	3.9	–	–
South Atlantic	120,227	21.1	104,154	25.5	51,827	18.6	–	–
East South Central	15,397	2.7	10,376	2.5	7,331	2.6	–	–
West South Central	165,419	29.1	110,441	27.1	83,226	29.9	–	–
Mountain	14,327	2.5	8,731	2.1	7,819	2.8	–	–
Pacific	15,302	2.7	10,361	2.5	7,089	2.6	–	–
International, unknown, or missing	105,645	18.6	69,507	17.0	56,022	20.2	–	–

6.2. Foundational Skills

6.2.1. Descriptive Statistics of Raw Scores

Appendix B presents the descriptive statistics of the Foundational Skills raw scores for Fall 2020 through Spring 2021. Sentence Reading Fluency is included in these tables but does not contribute to students' Foundational Skills domain scores. Because the Foundational Skills track is multistage adaptive, no student takes all the measures within a single testing session and the sample sizes for measures that appear later in the branching structure are smaller than those that appear near the beginning of the branching structure. Descriptive statistics for Sentence Reading Fluency are included even though the results do not contribute to students' Foundational Skills domain scores and nearly all passage-reading students also take this measure. Consequently, sample sizes for Sentence Reading Fluency are considerably larger than those for the Foundational Skills measures in first grade and above.

6.2.2. IRT Calibration of Foundational Skills Measures

Item response theory (IRT) is a statistical modeling technique that places items and persons onto the same scale in a manner that, given particular assumptions, is sample-independent. IRT allows student scores to be readily compared even when students have taken different sets of items on the same scale. For example, a student taking measures on the lower branches of the Foundational Skills lattices might have the same raw score as a student taking measures on the higher branches of the lattices, but the IRT ability estimate will be higher for the student taking the more difficult items. Separate Rasch IRT calibrations were conducted for each Foundational Skills domains: Phonological Awareness, Phonics & Word Recognition, and Language Comprehension.

6.2.2.1. Item Bank Construction

The IRT calibrations are intended to provide an item bank for future scoring of students, support the creation of scaled scores for each domain, and provide an historical, longitudinal dataset to support statistical modeling of student risk for reading difficulties, including dyslexia. The Rasch model was used for all item calibrations and student scoring. The model—which remains popular given its theoretical parsimony, ease of estimation, and usability, even with small samples—expresses the probability of a student of a particular ability providing a correct answer to an item of a particular difficulty (Equation 6.1; adapted from Wright & Stone, 1979):

$$P(\theta_i) = \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)} \quad (6.1)$$

where θ is the ability of student i, and b is the difficulty of item j.

NWEA psychometricians conducted calibrations on MAP Reading Fluency operational Foundational Skills measures for Fall 2020, Winter 2021, and Spring 2021. Spring 2021 was chosen as the reference or "bank" term because it was the latest term with the largest sample size (N > 438,000 students). All item calibrations were conducted with the Winsteps® Rasch measurement computer program (Linacre, 2021). Separate calibrations were conducted for each domain.

Table 6.3 presents the minimum and maximum number of responses for each item, the point-biserial correlations, and the weighted item fit statistics. The point-biserial correlation represents the correlation between how well students did on an item and how well they did on the domain. An item with a high positive point-biserial correlation distinguishes between low-performing and high-performing students better than an item with a point-biserial correlation near zero. The infit mean square statistic indicates how well the item difficulty estimate corresponds to the response pattern observed in the data. It has an expected value of 1.00 and is sensitive to unexpected responses when the item is targeted at respondents' ability level. The table also presents summary statistics for the item difficulty estimates. One item was removed from the bank for Language Comprehension for misfit to the Rasch model. The mean item difficulty estimate for Phonics & Word Recognition is offset from zero because the Sentence Reading Fluency measure was calibrated with the Phonics & Word Recognition measures but is not included in the item bank or student Phonics & Word Recognition ability estimates.

Table 6.3. Foundational Skills Item Bank Statistics—Spring 2021

		***N* Students**		**Point-Biserial Correlation**		**Infit Mean Square**		**Item Difficulty Estimates**			
Domain*	***N* Items**	**Min.**	**Max.**	**Min.**	**Max.**	**Min.**	**Max.**	**Mean**	**SD**	**Min.**	**Max.**
PA	277	8,282	58,609	0.11	0.61	0.83	1.32	0.00	1.19	-2.16	2.90
PWR	291	17,769	78,307	0.23	0.64	0.75	1.32	-0.12	1.32	-2.63	4.09
LC	79	63,659	97,476	0.30	0.56	0.83	1.41	-0.03	1.00	-1.62	2.97

*PA = Phonological Awareness. PWR = Phonics & Word Recognition. LC = Language Comprehension.

6.2.2.2. Assessing Item Drift

An assumption of IRT is that a student's probability of answer an item correctly depends only on the item's difficulty and the student's ability. It is customary in item banking to assess whether item difficulty has substantially changed over time. To assess item drift, the unanchored estimates from a particular term and domain were linked to the bank difficulty estimates via the mean-sigma method (Kolen & Brennan, 2014, Equations 2.1 – 2.4). Only items common to both terms being compared were used in these calculations.

$$Slope = \frac{\hat{\sigma}^2_{\hat{b}_{bank}}}{\hat{\sigma}^2_{\hat{b}_{new}}} \quad (6.2)$$

$$Intercept = \hat{\mu}_{\hat{b}_{bank}} - Slope * \hat{\mu}_{\hat{b}_{new}} \quad (6.3)$$

$$\hat{b}_{j_{eq}} = Slope * \hat{b}_{j_{new}} + Intercept \quad (6.4)$$

where $\hat{b}_{bank}$ represents the bank item difficulty estimates, $\hat{b}_{new}$ represents the item difficulty estimates from the comparison term, and $\hat{b}\,j_{eq}$ represents the item difficulty estimate for item *i* from the comparison term equated to the item bank, and $\hat{b}\,j_{new}$ is the original item difficulty estimate from the comparison term.

An item is considered drifted for a term if its difficulty differed from the bank difficulty by more than 0.30 logits, which is the industry standard for considering an item "drifted." Table 6.4 presents the number of items that drifted by an absolute value of > 0.30 logits by term.

Not all items that show drift are excluded from scoring or re-anchored. Any item that shows extreme drift is excluded from scoring. The same two items from the Language Comprehension domain showed severe drift over several terms through Fall 2019. Item writers confirmed that these items were originally defective but had been corrected by Winter 2020.

The size of the drift was small for most items in Phonological Awareness and Phonics & Word Recognition domains, but they were clustered within specific measures, namely in Initial Sound Fluency, Blending Phonemes, and Letter-Sound Fluency. Out of an abundance of caution, item difficulties for these measures were re-anchored in Fall 2018, Winter 2019, and Spring 2019 to the observed difficulty estimates obtained in Rasch calibrations where all other items were anchored to their bank difficulty estimates.

Table 6.4. Number of Drifted Items

	N* Drifted Items by Domain								
	Phonological Awareness			**Phonics & Word Recognition**			**Language Comprehension**		
Term	**Total**	**Excluded from Scoring**	**Re-anchored?**	**Total**	**Excluded from Scoring**	**Re-anchored?**	**Total**	**Excluded from Scoring**	**Re-anchored?**
Fall 2018	12	–	Yes	10	1	Yes	7	2	No
Winter 2019	18	1	Yes	–	–	No	4	2	No
Spring 2019	13	–	Yes	3	–	No	3	2	No
Fall 2019	1	–	No	4	–	No	4	1	No
Spring 2020	9	–	No	3	–	No	–	–	No
Fall 2020	–	–	No	3	–	No	–	–	No
Winter 2021	–	–	No	–	–	No	–	–	No
Spring 2021	–	–	No	–	–	No	–	–	No

*An item was considered drifted for a term if its difficulty differed from the bank difficulty by more than 0.30 logits. Some of the same items drifted in several terms.

6.2.2.3. Student Scoring

Item difficulty estimates from the established item bank and any re-anchored item sets from the item drift analyses were used to obtain maximum-likelihood ability estimates for each student by domain and term (Xue, 2020). Extreme scores were handled through fencing to assist the maximum-likelihood ability estimates. Namely, students obtaining a perfect minimum or maximum score were assigned two fictitious items, one very easy and scored as correct and the other very difficult and scored as incorrect. When adding a ghost item response for an extreme score, the optimizer of the ability estimation can find a boundary of the student ability and complete the estimation. Such fencing has no effect on a student's ability estimate and establishes a finite likelihood function for estimation (Han, 2016).

Standard errors for each ability estimate were estimated as the reciprocal of the square root of the test information function at that ability estimate. The test information function is the sum of the item information functions for the items presented to a student for a particular domain in a particular term. The item information function for the Rasch model is the probability of a student with a particular estimated ability answering an item correctly multiplied by the probability of answering the item incorrectly (Equations 6.5 – 6.7; adapted from Wright & Stone, 1979):

$$I_j = P_j\left(\hat{\theta}_i\right)\left[1 - P_j\left(\hat{\theta}_i\right)\right], \tag{6.5}$$

$$I_T = \sum_{j=1}^{m} I_j, \tag{6.6}$$

$$S_\theta = \frac{1}{\sqrt{I_T}}, \tag{6.7}$$

where I_j is the information function for item *j*, $P_j\left(\hat{\theta}_i\right)$ is the probability of a correct response to item *j* from person *i* with estimated ability, and $\hat{\theta}$, I_T is the test information function for test *T*.

6.2.2.4. Sample Refinement

Students not attempting at a total of at least 10 calibrated items spanning at least two measures within a domain were excluded from further analysis involving the ability estimates for that domain. Fewer than 1% of the student records in any domain were removed.

6.2.3. Marginal Reliability

Marginal reliability is an IRT-based technique to estimate the reliability of a test (Samejima, 1977, 1994). The calculations are based on the definition of reliability as the proportion of total variance that is considered true score variance (Equation 6.8). The standard errors of individual students are averaged across observations within the reporting group (e.g., grade, term).

$$\rho_\theta = \frac{\hat{\sigma}_\theta^2 - \hat{\mu}_{S_\theta^2}}{\hat{\sigma}_\theta^2} \tag{6.8}$$

where $\hat{\sigma}_{\theta}^{2}$ is the observed variance of the ability estimates, θ, and $\hat{\mu}_{S_{\theta}^{2}}$ is the observed mean of the score's conditional error variances at each value of θ. Tests are considered of sound reliability when their marginal reliability coefficients range from 0.80 and above according to the National Center on Intensive Intervention (NCII, 2020).

Table 6.5 presents the median marginal reliability coefficient with a 95% confidence interval for 1,000 bootstrapped samples stratified on student grade and term. The bootstrap sampling method allows inference of the population marginal reliability coefficient from the observed data. Reliabilities are reported when 300 or more student records were available for a particular term and grade combination. Any result with *N* < 300 is indicated by † in the table.

The lower limit of the confidence intervals exceeded 0.80 for Phonological Awareness and 0.90 for Phonics & Word Recognition in nearly all grades and terms. Reliabilities were lower for Language Comprehension, which may be because of the Language Comprehension measure is much shorter. Language Comprehension also suffers from ceiling effects. For example, approximately 25% of the Grade 4 and Grade 5 students answered all items correctly in this domain. NWEA content staff have created a set of more difficult Language Comprehension items that will be added to the measure in a future school year.

Table 6.5. Bootstrapped Marginal Reliability Coefficients

	Fall 2020				Winter 2021				Spring 2021			
			95% CI				95% CI				95% CI	
Grade	*N*	Median	Lower	Upper	*N*	Median	Lower	Upper	*N*	Median	Lower	Upper
Phonological Awareness												
K	75,412	0.904	0.904	0.905	70,260	0.911	0.911	0.912	95,166	0.913	0.912	0.913
1	91,393	0.892	0.891	0.893	72,632	0.892	0.891	0.893	73,659	0.890	0.889	0.891
2	61,536	0.858	0.856	0.859	35,068	0.870	0.868	0.872	32,973	0.877	0.875	0.879
3	13,442	0.847	0.843	0.851	9,175	0.850	0.845	0.854	7,674	0.863	0.859	0.867
4	844	0.855	0.840	0.867	547	0.850	0.832	0.866	503	0.851	0.834	0.867
5	337	0.870	0.844	0.889	320	0.813	0.779	0.841	†	†	†	†
Phonics & Word Recognition												
K	75,533	0.940	0.940	0.941	70,398	0.940	0.940	0.941	95,325	0.945	0.945	0.945
1	91,674	0.936	0.935	0.936	72,794	0.935	0.934	0.936	73,881	0.941	0.940	0.941
2	61,760	0.926	0.925	0.927	35,239	0.928	0.927	0.929	33,122	0.932	0.931	0.933
3	13,532	0.912	0.910	0.914	9,241	0.914	0.912	0.917	7,718	0.921	0.917	0.923
4	847	0.915	0.905	0.924	546	0.914	0.902	0.925	503	0.912	0.901	0.923
5	338	0.913	0.895	0.927	322	0.885	0.855	0.910	†	†	†	†
Language Comprehension												
K	76,791	0.838	0.837	0.840	70,845	0.812	0.811	0.814	95,646	0.794	0.792	0.795
1	91,925	0.761	0.759	0.763	72,929	0.740	0.737	0.742	73,998	0.731	0.728	0.734
2	61,846	0.689	0.685	0.692	35,297	0.731	0.727	0.735	33,184	0.745	0.740	0.749
3	13,557	0.711	0.703	0.719	9,270	0.717	0.707	0.726	7,741	0.740	0.731	0.750
4	847	0.566	0.515	0.613	548	0.558	0.485	0.617	503	0.609	0.550	0.656
5	338	0.568	0.481	0.634	322	0.479	0.369	0.568	†	†	†	†

† N < 300

6.2.4. Establishment of Score Scale

Student ability estimates in the theta metric were transformed to the scaled score metric using the Winter 2020 data as the reference term based on its large sample across the available data at the time of the score scale establishment. This process converts the student ability estimates onto a common score scale, which allows for legitimate and meaningful comparisons across test forms, school terms, school years or grades. The chosen score scale has a mean of 500 and a standard deviation of 10. For each domain, the slope and intercept for the transformation were calculated as Equations 6.9 and 6.10. Table 6.6 presents the slope and intercept for each domain.

$$Slope = \frac{10}{\hat{\sigma}_{\theta}^{2}} \tag{6.9}$$

$$Intercept = 50 - Slope * \hat{\mu}_{\theta} \tag{6.10}$$

Table 6.6. Slopes and Intercepts for Scaled Score Transformation

Domain	Slope	Intercept
Phonological Awareness	6.303130	494.093595
Phonics & Word Recognition	5.645238	494.872744
Language Comprehension	6.426431	485.928829

The slopes and intercepts were then used to transform student theta estimates to scaled scores (*SS*) for all terms:

$$SS = Slope * \hat{\theta} + Intercept \tag{6.11}$$

The slopes were used to transform theta standard errors to standard errors of measurement (SEMs) of the scaled scores:

$$SEM = Slope * SE\left(\hat{\theta}\right) \tag{6.12}$$

Given that the theta estimation of MAP Reading Fluency Foundational Skills currently runs from a possible low of -10 to a possible maximum of 10, the corresponding reporting intervals for the scaled domain scores are shown in Table 6.7.

Table 6.7. Reporting Intervals for Scaled Foundational Skills Domain Scores

Domain	Reporting Interval
Phonological Awareness	[431, 557]
Phonics & Word Recognition	[438, 551]
Language Comprehension	[422, 550]

Table 6.8 presents descriptive statistics for the scaled scores and SEMs. The scaled score means display the desirable property of mostly increasing with term and grade. Grade 3 and above tend to involve intervention populations, so a strict increase in scaled score means is not necessarily expected. These scores will become available for the MAP Reading Fluency assessments in Fall 2022.

Table 6.8. Descriptive Statistics for Scaled Scores and SEMs—Foundational Skills

	Fall 2020					Winter 2021					Spring 2021				
		SS		SEM			SS		SEM			SS		SEM	
Grade	*N*	Mean	SD	Mean	SD	*N*	Mean	SD	Mean	SD	*N*	Mean	SD	Mean	SD
Phonological Awareness															
K	75,412	490.94	9.25	2.70	0.93	70,260	494.36	9.44	2.67	0.88	95,166	497.38	9.87	2.76	0.95
1	91,393	498.90	9.16	2.84	1.01	72,632	501.34	9.58	2.95	1.10	73,659	502.42	9.79	3.02	1.18
2	61,536	503.57	8.94	3.14	1.21	35,068	503.53	9.32	3.13	1.23	32,973	503.24	9.56	3.11	1.24
3	13,442	504.10	8.89	3.24	1.27	9,175	504.93	9.14	3.29	1.32	7,674	504.72	9.59	3.28	1.35
4	844	506.62	9.19	3.27	1.26	547	507.42	9.40	3.36	1.41	503	507.93	9.93	3.50	1.53
5	337	507.51	10.01	3.35	1.31	320	508.26	8.71	3.44	1.49	237	508.36	10.22	3.41	1.41
Phonics & Word Recognition															
K	75,533	489.86	10.46	2.38	0.94	70,398	493.51	9.57	2.20	0.79	95,325	496.71	9.76	2.17	0.72
1	91,674	499.21	9.36	2.24	0.78	72,794	500.84	8.98	2.17	0.72	73,881	502.25	9.56	2.19	0.77
2	61,760	505.33	9.38	2.38	0.94	35,239	504.15	8.94	2.25	0.84	33,122	503.72	9.04	2.20	0.84
3	13,532	506.21	8.70	2.39	0.98	9,241	506.54	8.90	2.40	1.01	7,718	506.23	9.11	2.36	1.02
4	847	509.88	9.96	2.64	1.19	546	510.34	10.09	2.69	1.22	503	510.37	10.27	2.72	1.35
5	338	512.02	10.47	2.79	1.30	322	511.09	8.97	2.71	1.35	235	510.75	11.02	2.62	1.20
Language Comprehension															
K	76,791	493.05	8.88	3.37	1.18	70,845	495.77	9.20	3.70	1.46	95,646	497.63	9.59	4.01	1.69
1	91,925	499.17	9.43	4.24	1.81	72,929	500.70	9.74	4.56	1.98	73,998	501.48	9.99	4.74	2.08
2	61,846	502.65	9.63	4.94	2.12	35,297	501.94	10.25	4.87	2.13	33,184	501.69	10.47	4.84	2.14
3	13,557	502.83	10.24	5.06	2.18	9,270	502.95	10.49	5.11	2.23	7,741	502.60	10.92	5.09	2.23
4	847	506.52	9.62	5.86	2.37	548	507.26	9.85	6.09	2.40	503	506.03	9.99	5.76	2.40
5	338	507.86	10.37	6.34	2.49	322	507.82	9.14	6.17	2.26	237	507.43	10.46	6.20	2.54

6.2.5. Validity Studies

Concurrent validity is expressed in the form of a Pearson correlation coefficient between a test score and the score of another established and validated test designed to assess the same content. It answers the question, “How well do the scores from this test that reference a particular scale correspond to the scores obtained from another test that references some other scale in the same subject?” Concurrent validity requires that both tests are administered to the same students within a short amount of time. Concurrent evidence is a typical part of many validity arguments, as two tests of the same construct should share a strong statistical relationship.

Table 6.9 presents the correlations between the Foundational Skills domain scores and MAP Growth Reading scores by grade and term. Most correlations for Phonological Awareness and Phonics & Word Recognition are in the 0.60s and 0.70s. Considering that the correlations reflect a part-whole relationship (i.e., a portion of the Foundational Skills assessment correlated to the overall reading achievement in MAP Growth Reading), they appear adequate. Correlations are lower for Language Comprehension in Grade 3, which is likely due to the ceiling effects in the Language Comprehension measures.

Table 6.9. Correlations between Foundational Skills Domain Scores and MAP Growth Reading Scores (Concurrent Validity)—2020–2022 School Year

Term	Grade	Phonological Awareness		Phonics & Word Recognition		Language Comprehension	
		N	*r*	N	*r*	N	*r*
Fall 2020-2022	K	23,342	0.54	23,342	0.63	23,342	0.53
	1	38,930	0.66	38,930	0.71	38,930	0.59
	2	12,803	0.64	12,803	0.71	12,803	0.57
	3	11,977	0.54	11,977	0.71	11,977	0.42
Winter 2020-2022	K	29,882	0.65	29,882	0.68	29,882	0.60
	1	31,487	0.68	31,487	0.71	31,487	0.60
	2	8,563	0.64	8,563	0.71	8,563	0.56
	3	7,844	0.55	7,844	0.71	7,844	0.43
Spring 2020-2022	K	28,651	0.69	28,651	0.69	28,651	0.60
	1	28,544	0.67	28,544	0.71	28,544	0.57
	2	7,786	0.66	7,786	0.72	7,786	0.56
	3	5,420	0.54	5,420	0.69	5,420	0.41

6.3. Oral Reading Fluency

6.3.1. Descriptive Statistics of Raw WCPM Scores

Table 6.10 presents the descriptive statistics for students' average raw WCPM scores for the 2020–2021 school year.[6] N-counts in the oral reading fluency sample (Table 6.1) are larger than those in this table because not all students who begin the Oral Reading Fluency track finish it or produce a machine-scoreable response. As expected, scores largely increased with grade. Grade K scores sometimes exceeded those for Grade 1, which may be because kindergarteners who progress to the Oral Reading fluency track are likely higher-ability readers to begin with.

Table 6.10. Descriptive Statistics of Average WCPM Scores

	Fall 2020			Winter 2021			Spring 2021		
Grade*	N	Mean	SD	N	Mean	SD	N	Mean	SD
K	838	99.76	39.11	1,332	85.03	34.67	3,071	76.29	29.07
1	6,107	81.62	28.22	11,207	76.90	26.23	23,802	79.15	26.59
2	23,774	84.02	26.96	24,174	83.24	26.07	40,871	89.04	28.23
3	22,289	87.82	26.70	18,488	93.26	26.76	22,304	96.58	28.28
4	11,209	89.74	28.88	9,941	95.59	28.77	11,414	100.18	29.71
5	9,031	99.50	29.66	6,668	100.21	28.79	7,098	105.13	30.07
6	821	109.65	32.05	786	106.22	29.91	850	111.56	31.42
7	448	101.96	29.48	148	98.64	28.85	135	117.83	30.83
8	172	118.82	31.99	86	114.26	34.36	81	119.90	33.90
9	–	–	–	†	†	†	–	–	–
10	†	†	†	†	†	†	†	†	†

*Pre-K is not included because there are no expectations of oral reading fluency until spring of Grade 1. Kindergarteners are included because there are a fair number of Grade K passage readers in the spring.
† N < 25

Table 6.11 presents the descriptive statistics for the Passage Comprehension Quiz questions that follow a student's reading of a passage. Raw scores can range from 0–12. A student who reads one passage can have a maximum raw score of six, and a student who reads two passages can have a maximum raw score of 12. A passing score is 5–6 questions correct for a passage (i.e., 80% correct). On average, students scored near or above this criterion by fall or winter of Grade 1. There is not a strict increase in either raw or proportion correct scores across grades or terms. Given that the Passage Comprehension Quiz scores are not scaled or equated, variance in item difficulty could be affecting this progression of means.

[6] Field test passages were not included in a student's average WCPM score.

Table 6.11. Descriptive Statistics for Passage Comprehension Quiz

	Fall 2020				Winter 2021				Spring 2021			
		Raw Scores		Proportion Correct		Raw Scores		Proportion Correct		Raw Scores		Proportion Correct
Grade	N	Mean	SD	Mean	N	Mean	SD	Mean	N	Mean	SD	Mean
K	838	9.14	2.18	0.77	1,331	9.42	2.22	0.79	3,071	9.30	2.05	0.78
1	6,107	10.10	1.73	0.84	11,206	10.09	1.89	0.84	23,801	9.89	1.86	0.83
2	23,774	9.97	1.93	0.83	24,150	9.86	2.00	0.82	40,827	9.62	1.95	0.80
3	22,289	9.80	1.72	0.82	18,468	10.37	1.67	0.87	22,260	9.84	1.77	0.82
4	11,209	9.46	1.93	0.79	9,941	10.28	1.62	0.86	11,412	9.82	1.90	0.82
5	9,031	9.83	2.13	0.82	6,668	10.30	1.65	0.86	7,097	9.96	1.98	0.83
6	821	10.10	2.11	0.84	786	10.20	1.73	0.85	850	9.88	2.07	0.82
7	448	9.37	2.29	0.78	148	9.95	1.99	0.83	135	9.90	2.15	0.83
8	172	9.68	2.28	0.81	86	10.34	1.61	0.86	81	9.91	1.81	0.83
9	–	–	–	–	†	†	†	†	–	–	–	–
10	†	†	†	†	–	–	–	–	†	†	†	†

† N < 25

6.3.2. Passage Equating

Equated WCPM scores were introduced in Fall 2019. Equating is a statistical procedure that makes the WCPM scores of different passages comparable. Since the WCPM scores are affected by passage difficulty, they are equated to SWCPM scores so that the equated scores can be used interchangeably and can be compared across students. Each passage was equated to an anchor passage to allow the conversion of WCPM scores to equated scores. The first round of equating was conducted with data from 2018–2019. A second round, using data from Fall 2019 and Winter 2020, was later conducted for field test and other previously unequated passages. These two rounds of equating accounted for nearly all operational passages. Equating in both rounds employed a single-groups design. The "equate" package for R was used for all conversions. Equipercentile equating with loglinear pre-smoothing preserving two moments was used for the 2018–2019 data. Linear equating was used for the 2019–2020 data.

For the 2018–2019 data, a passage of medium difficulty from the Winter 2019 data was chosen as the main anchor passage. Other passages from Winter 2019 were equated to scores on this anchor passages. Chained equipercentile equating was then used to equate passages from Fall 2018 and Spring 2019 to the Winter 2019 anchor passage. For each conversion, students whose anchor-target passage pair scores showed a squared Mahalanobis distance ≥ 10 were removed from the equating sample (Equation 6.13). The Mahalanobis distance was chosen as the statistic to identify outliers because it accounts for the covariance between the two sets of scores. Fewer than 2% of the sample was removed for any anchor-target passage pair.

$$D_{mh}^{2} = (x - \mu_x)\Sigma^{-1}(y - \mu_y) \quad (6.13)$$

One evaluation measure used for the 2018–2019 school year equatings was the reduction in within-student variance. Table 6.12 presents the average within-student variance for raw WCPM scores versus SWCPM scores. These values, on average, represent a 40–61% reduction in within-student variance. Reduced within-student variance suggests that the equating has successfully controlled for passage difficulty effects.

Table 6.12. SD of Raw and Scaled WCPM Scores, 2018–2019 School Year

Term	Average Within-Student SD	
	Raw WCPM	Scaled WCPM
Fall 2018	12.19	7.59
Winter 2019	9.52	6.94
Spring 2019	9.92	7.69

Root mean square deviations (RMSDs) between students' scores on anchor passages and equated passages were also created. RMSDs ranged from 11.50 to 13.69. These are slightly higher than desired but adequate. For the Fall 2019 and Winter 2020 data, two previously equated passages were chosen as anchors. The Spring 2020 data did not offer enough observations for equating for any passage pair. The research team opted for stricter outlier removal compared with that of the 2018–2019 school year. For each conversion, students whose anchor-target passage pair scores showed a squared Mahalanobis distance of ≥ 5 were removed from the equating sample for that conversion. A maximum of 9.1% of the records were removed for any conversion. Sample sizes in the final equating sample ranged from 974 to 2,298 across anchor target passage pairs. Correlations between WCPM scores ranged from 0.85 to 0.91. Liu and Walker (2011) recommend correlations > 0.866 for to-be-equated scores and their anchor scores. Of the 48 target passages for the 2019–2020 school year, only three had anchor-target correlations < 0.87. RMSDs between anchor- and equated-passage scores ranged from 8.94 to 12.02 WCPM. These RMSDs are also larger than desired but adequate.

One thousand replications were conducted for each conversion to obtain the bootstrapped standard error of estimate (SEE) for these conversions. The SEEs were small throughout the 20–200 WCPM range. A WCPM of 200 would fall slightly below the 90th percentile for spring Grade 6 in the Hasbrouck and Tindal (2017) norms. Most SEEs in this WCPM range were < 2 WCPM, and the maximum was 3.68 WCPM.

New passages are currently field tested as student's third passage if they receive one. These new passages are equated to the reference passage at the end of each school year. Table 6.13 presents the descriptive statistics for students' average SWCPM by grade and term for the 2020–2021 school year.

Table 6.13. Descriptive Statistics of Average SWCPM

Grade	Fall 2020			Winter 2021			Spring 2021		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
K	838	96.47	35.28	1,332	83.49	33.23	3,071	78.26	27.99
1	6,107	79.41	25.54	11,207	76.27	25.74	23,802	81.45	25.76
2	23,774	84.20	24.50	24,174	87.48	24.80	40,871	94.53	26.43
3	22,289	96.52	26.17	18,488	100.86	25.78	22,304	106.82	26.63
4	11,209	103.02	28.40	9,941	108.64	28.27	11,414	115.80	28.59
5	9,031	118.25	28.50	6,668	118.39	28.58	7,098	125.51	28.99
6	821	127.34	29.15	786	124.30	29.22	850	130.91	28.79
7	448	120.27	27.37	148	116.76	29.82	135	137.33	28.29
8	172	134.89	27.67	86	130.90	33.36	81	137.44	30.86
9	–	–	–	†	†	†	–	–	–
10	†	†	†	–	–	–	†	†	†

† N < 25

6.3.3. Test-Retest Reliability

Test-retest reliability measures the correlation between two test events for the same students and provides insight into the consistency of the MAP Reading Fluency construct across time. Tests are considered of sound reliability when their test-retest reliability coefficients range from 0.70 and above. To calculate test-retest reliability, students with more than two test events were selected and their first two records were subset in chronological order. From this subset, students were excluded whose two test events were less than one day or more than 31 days apart. Pearson correlation coefficient between the average WCPM for operational passages of these two test events were then computed across all valid students. As shown in Table 6.14, test-retest reliabilities are all higher than 0.80, suggesting that the MAP Reading Fluency oral reading scores show consistency of measurement.

Table 6.14. Test-Retest Reliability for Average Oral Reading Fluency Scores—2018–2019 School Year

	Fall 2018		Winter 2019		Spring 2019	
Grade	N	*r*	N	*r*	N	*r*
1	69	0.89	233	0.89	364	0.81
2	389	0.83	533	0.86	783	0.88
3	488	0.88	337	0.91	364	0.89

6.3.4. Validity Studies

6.3.4.1. 2019 Human-Machine Agreement

NWEA commissioned a human-machine agreement study in 2019–2020. When the first human-machine reliability study was conducted in 2017, MAP Reading Fluency had only 10 passages. At the time of this study, there were more than 170 English passages at the time of this study, which necessitated a new passage scoring study in collaboration LanguaMetrics. Strategic Measurement and Evaluation, Inc. (SME) provided professional human raters, jointly trained by SME and NWEA.

Stratified sampling was applied to randomly selected passage records from Fall 2019 to represent student demographics. English language learners (ELLs) were oversampled to ensure adequate representation and sample size. The final sample contained 1,728 responses from 108 passages, of which 476 were sampled from ELL students. Only machine-scorable records were included as valid responses. Table 6.15 presents the sample demographics. Invalid passage records were rejected due to reasons such as distortion and background noise and were excluded from further analysis. The final data set contained 1,362 records.

SME raters hand-scored the passage reading responses from the sample. Each passage was rated by at least two raters. When primary raters disagreed on WCPM scores by 3 or more points, an additional score was applied by a master rater. A consensus decision was then determined by the majority. For example, if two out of three raters scored a word as error, the consensus was determined as error.

Table 6.15. Sample Demographics for Human-Machine Agreement—Fall 2019

Demographic Subgroup	*N* Students																		Total	
	Grade K		Grade 1		Grade 2		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8			
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Sex																				
Male	8	0.47	129	7.52	262	15.27	290	16.90	112	6.53	56	3.26	9	0.52	2	0.12	–	–	868	50.58
Female	12	0.70	113	6.59	245	14.28	263	15.33	125	7.28	70	4.08	11	0.64	6	0.35	3	0.17	848	49.42
Total	20	1.17	242	14.10	507	29.55	553	32.23	237	13.81	126	7.34	20	1.17	8	0.47	3	0.17	1,716	100.00
																		#Missing	12	–
Race/Ethnicity*																				
AI/AN	–	–	1	0.06	8	0.46	5	0.29	3	0.17	1	0.06	–	–	1	0.06	1	0.06	20	1.16
Asian/PI	–	–	32	1.86	29	1.68	17	0.99	1	0.06	4	0.23	–	–	–	–	–	–	83	4.81
Black or African American	1	0.06	33	1.91	69	4.00	82	4.76	38	2.20	22	1.28	3	0.17	1	0.06	–	–	249	14.44
Hispanic or Latino	3	0.17	18	1.04	75	4.35	229	13.28	24	1.39	23	1.33	–	–	5	0.29	1	0.06	378	21.93
Multi-Ethnic	–	–	7	0.41	19	1.10	23	1.33	6	0.35	2	0.12	2	0.12	–	–	–	–	59	3.42
NH/PI	–	–	–	–	2	0.12	–	–	1	0.06	–	–	–	–	–	–	–	–	3	0.17
Not Specified/Other	3	0.17	25	1.45	36	2.09	32	1.86	18	1.04	7	0.41	1	0.06	1	0.06	–	–	123	7.13
White	13	0.75	129	7.48	274	15.89	165	9.57	146	8.47	67	3.89	14	0.81	–	–	1	0.06	809	46.93
Total	20	1.16	245	14.21	512	29.70	553	32.08	237	13.75	126	7.31	20	1.16	8	0.46	3	0.17	1,724	100.00
																		#Missing	4	–

*AI/AN = American Indian/Alaskan Native. PI = Pacific Islander. NH/PI = Native Hawaiian/Other Pacific Islander.

The scored data file contains raters' consensus WCPM, raters' prosody score, and machine WCPM score. Table 6.16 presents the descriptive statistics of the human consensus vs. machine WCPM.

Table 6.16. Descriptive Statistics for Human vs. Machine WCPM—Fall 2019

Demographic Subgroup		*N*	Mean	SD	Min.	Max.
Overall	Human Consensus WCPM	1,362	79.23	29.82	16.00	210.00
	Machine WCPM		78.76	29.22	18.00	212.00
ELL	Human Consensus WCPM	384	80.77	26.26	28.00	176.00
	Machine WCPM		80.29	25.71	28.00	179.00
Sex						
Female	Human Consensus WCPM	703	79.65	30.14	22.00	177.00
	Machine WCPM		79.55	29.51	22.00	181.00
Male	Human Consensus Score	651	79.00	29.35	16.00	210.00
	Machine Score		78.14	28.77	18.00	212.00
Race/Ethnicity						
American Indian/Alaskan Native	Human Consensus Score	17	86.82	36.26	25.00	144.00
	Machine Score		87.47	36.76	28.00	145.00
Asian/Pacific Islander	Human Consensus Score	72	80.31	29.82	33.00	168.00
	Machine Score		79.68	28.89	33.00	165.00
Black or African American	Human Consensus Score	188	77.37	27.32	27.00	169.00
	Machine Score		76.41	26.17	29.00	163.00
Hispanic or Latino	Human Consensus Score	303	80.97	25.83	29.00	174.00
	Machine Score		80.17	25.29	28.00	179.00
Multi-Ethnic	Human Consensus Score	47	80.98	34.84	19.00	176.00
	Machine Score		80.60	34.27	20.00	175.00
Native Hawaiian/Other Pacific Islander	Human Consensus Score	1	104.00	N/A	104.00	104.00
	Machine Score		96.00	N/A	96.00	96.00
Not Specified/Other	Human Consensus Score	85	80.75	33.10	26.00	167.00
	Machine Score		81.07	32.78	26.00	170.00
White	Human Consensus Score	649	78.27	31.25	16.00	210.00
	Machine Score		78.00	30.71	18.00	212.00

The following methods, which are common in evaluating scoring consistency, were used to evaluate the results:

1. Root-mean-square-difference (RMSD): measures the differences between human and machine scores:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(human_i - machine_i)^2}{N}} \quad (6.14)$$

2. Pearson correlation (*r*): The correlation between human and machine scores
3. Proportion agreement: Proportion of decisions on which human consensus rating and machine scores agreed

Table 6.17 presents these statistics for the Fall 2019 data. The Pearson correlations are all above 0.98, the RMSDs are small, and the proportion agreement is high, suggesting that the human and machine scores are highly consistent overall and across subgroups.

Table 6.17. Human-Machine Agreement

Demographic Subgroup	*N*	RMSD	*r*	Proportion Agreement
Overall	1,362	4.79	0.99	0.99
ELL	384	4.75	0.98	0.99
Female	703	4.02	0.99	1.00
Male	651	5.51	0.98	0.99
American Indian/Alaskan Native	17	2.26	1.00	1.00
Asian/Pacific Islander	72	5.55	0.98	0.99
Black or African American	188	5.44	0.98	0.99
Hispanic or Latino	303	4.94	0.98	0.99
Multi-Ethnic	47	3.43	1.00	1.00
Native Hawaiian/Other Pacific Islander	1	†	†	†
Not Specified/Other	85	3.83	0.99	1.00
White	649	4.66	0.99	0.99

† N < 25

6.3.4.2. 2019 Concurrent Validity

MAP Reading Fluency scores should show strong statistical relationships with scores from well-established tests of early reading. The DIBELS® family of products is a long-established and respected set of reading fluency tests. Many testing programs validate the use of scores from their reading fluency assessments against assessments from the DIBELS/Acadience family. (Good et al., 2011, 2013–2019; University of Oregon, 2018–2020).

NWEA conducted comparisons of DIBELS Next/Acadience oral reading fluency scores with those from MAP Reading Fluency using data from the 2018–2019 school year. Partner districts received financial incentives to provide their DIBELS data. After merging these DIBELS Next/Acadience records with the corresponding MAP Reading Fluency scores, the result was data of 622 students from nine school districts across four geographic census divisions (East North Central, Middle Atlantic, Pacific, and West North Central). The scores of interest were the DIBELS Next/Acadience Oral Reading Fluency score and each student's average SWCPM score from MAP Reading Fluency.

Table 6.18 presents the sample sizes and correlations of DIBELS Next/Acadience oral reading scores with MAP Reading Fluency's SWCPM scores. Neither DIBELS nor MAP Reading Fluency has expectations for oral reading fluency performance until Winter of Grade 1. All correlations exceeded 0.70, most were ≥ 0.85, and one was > 0.90. These correlations provide excellent evidence consistent with both oral reading fluency assessments measuring the same construct.

Table 6.18. Correlations of DIBELS Next/Acadience and MAP Reading Fluency Oral Reading Fluency Scores—2018–2019 School Year

	DIBELS Next/Acadience Correlations with Average MAP Reading Fluency SWCPM Scores					
	Fall 2018		Winter 2019		Spring 2020	
Grade	N	*r*	N	*r*	N	*r*
K	–	–	–	–	–	–
1	–	–	99	0.84	79	0.87
2	70	0.90	209	0.87	190	0.89
3	35	0.74	35	0.85	33	0.74

6.4. Effectiveness of Sentence Reading Fluency in Classifying Oral Reading Fluency

Sentence Reading Fluency is the routing test for all Adaptive Oral Reading forms except the Grade 4+ form. Students always proceed to passage reading after the Sentence Reading Fluency measure. Students meeting or exceeding a raw score of 15 and obtaining at least 75% of attempted items correct proceed on to passage reading. A cut point of 30 WCPM was selected for successful independent passage reading. This cut point is slightly above the 50th percentile for winter Grade 1 in the Hasbrouck and Tindal (2017) norms. Results from preliminary receiver operating characteristic (ROC) curve analyses and expert judgement were used to determine the cut points for Sentence Reading Fluency and the WCPM scores.

Correlations suggested that Sentence Reading Fluency would be an excellent predictor of oral reading fluency scores for most grades. The correlations between Sentence Reading Fluency and WCPM scores for individual passages ranged from 0.72 to 0.91 for the Grades K–2 passages and from 0.62 to 0.64 for the Grade 3 passages (NWEA, 2019). NWEA researchers investigated the classification accuracy of these cut points in Sentence Reading Fluency regarding performance in oral reading fluency

Logistic regressions with above/below cut point status on Sentence Reading Fluency as the predictor variable and above/below cut point status on WCPM scores for reading passages were conducted. Separate regressions were conducted for each passage, and only the raw score cut point could be used for Sentence Reading Fluency.

For binary responses y (y=1 or y=0), the linear logistic model is:

$$\hat{p}(y=1) = \frac{exp(a+\mathbf{BX})}{1+exp(a+\mathbf{BX})}, \tag{6.15}$$

where $\hat{p}$ is the predicted probability of being in the group labeled “1”, a is the intercept, and $\mathbf{BX}$ is a vector of regression weights and predictor scores.

The terms used for the classification accuracy results included sensitivity, specificity, false positive rate, false negative rate, base rate, and overall classification accuracy, as shown in Figure 6.1.

Figure 6.1. Classification of Oral Reading Fluency by Sentence Reading Fluency

		Observed Performance of Oral Reading Fluency		
		Above	Below	Total
Predicted Performance by Sentence Reading Fluency	Above	True Positive (TP)	False Positive (FP)	TP+FP
	Below	False Negative (FN)	True Negative (TN)	FN+TN
	Total	TP+FN	FP+TN	TP+FP+FN+TN

FPR=False Positive Rate=[FP/(FP + TN)]

FNR=False Negative Rate=[FN/(TP + FN)]

SEN=Sensitivity=[TP/(TP + FN)]

SEP=Specificity=[TN/(TN + FP)]

BR=Base Rate=[(TP+FN)/(TP+FP+FN+TN)]

OCR=Overall Classification Rate=[(TP+TN)/(TP+FP+FN+TN)]

Table 6.19 presents the classification accuracy statistics of MAP Reading Fluency across passages and grades. The sensitivity refers to the proportion of observations of good performance in oral reading fluency that are accurately identified as good by Sentence Reading Fluency. The specificity refers to the proportion of observations of poor oral reading fluency that are accurately identified as poor by sentence reading fluency. These two values often have an inverse relationship. Sensitivity is excellent for all grades except kindergarten. Specificity is less good for some passages in Grade 2 and Grade 3.

Table 6.19. Classification Accuracy Statistics for Oral Reading Passages—Winter 2017

Grade	Passage Title	Passage Code	Cutoff Value of Screening Test	FPR*	FNR*	SEN*	SPE*	BR*	OCA*
K	Sal Gets Wet	0111	13	0.00	0.36	0.64	1.00	0.67	0.76
	Pink the Pig	0112	13	0.00	0.35	0.65	1.00	0.65	0.77
1	Bears	1111	15	0.00	0.14	0.86	1.00	0.92	0.87
	Losing Teeth	1112	15	0.00	0.06	0.94	1.00	0.97	0.94
2	Old Photos	2111	15	0.25	0.03	0.97	0.75	0.98	0.96
	Game Inventor	2112	15	0.33	0.03	0.97	0.67	0.99	0.97
	Butterflies and Moths	2113	15	0.43	0.03	0.97	0.57	0.95	0.95
3	Bad Talent Show	3111	15	0.33	0.02	0.98	0.67	0.99	0.98
	Field Mice	3112	15	0.00	0.00	1.00	1.00	0.99	1.00
	Hamster on the Loose	3113	15	0.50	0.02	0.98	0.50	0.99	0.98

*FPR = false positive rate. FNR = false negative rate. SEN = sensitivity. SEP = specificity. BR = base rate. OCA = overall classification accuracy.

7. Dyslexia Screener

The MAP Reading Fluency Dyslexia Screener was first launched in March 2021, becoming available to all users in Fall 2021. The screener assesses key foundational reading skills, including those most often associated with dyslexia; applies a predictive model to flag student results that suggest possible risk factors for dyslexia or other reading difficulties; and provides actionable data to inform instruction and drills down into each student's strengths and needs. A flag on these reports is not a diagnosis of dyslexia or of a reading disability; rather, it is an indicator that the student's performance suggests possible risk factors for dyslexia or other reading difficulty.

More specifically, the Sentence Reading Fluency score and the Foundational Skills domain scores in Phonological Awareness, Phonics & Word Recognition, and Language Comprehension are included in a multivariate predictive model that flags student results, suggesting possible risk factors for dyslexia or other reading difficulties. Within the Phonological Awareness domain, skills with both larger phonological units and individual phonemes are assessed. Within the Phonics & Word Recognition domain, measures assess letter-sound knowledge, letter naming, word level decoding and word fluency skills, and word level encoding skills. The Language Comprehension domain measures oral language skills at the word and sentence levels. Rapid Automatized Naming (RAN) scores supplement the multivariate predictive model.

As shown in the theory of action (Jiban & Simpson, 2021), dyslexia screening data from MAP Reading Fluency support improved outcomes in three broad ways:

1. Students *flagged* as at increased risk are flagged by a multivariate predictive model. The recommendation is to consider these flagged students for *more resource-intensive follow ups* such as increased assessment, increased intensity of instruction, and increased communication with families. For students ready to read from passages, *progress monitoring* offers a faster feedback cycle for adjusting interventions.
2. MAP Reading Fluency reports support greater *data-based differentiation* for all students. In both the Phonological Awareness and Phonics & Word Recognition domains, students are pointed to research-based instructional activities aligned to their ZPD level.
3. *Instructional time* is returned via the efficiency of automatic, adaptive, group-administered screening of all students. By improving the foundational decoding skills that support reading fluency, students' long term passage comprehension outcomes are supported.

7.1. Dyslexia Screener Background

Research shows that early identification and intervention is highly effective in improving long-term reading outcomes. The earlier we can intervene with students likely to struggle, the more effective we can be. According to the International Dyslexia Association (IDA, n.d.), perhaps as much as 15–20% of the population exhibits characteristics of dyslexia, including slow or inaccurate word recognition, poor spelling, and difficulties with decoding. The IDA notes that in the school population nationwide, a significant number of students receiving services for learning disabilities have dyslexia. Many students with dyslexia are also currently unidentified and receive no services. While often diagnosed much later, signs of dyslexia may be evident as early as kindergarten. Dyslexia screening is therefore an invaluable tool for targeting early intervention.

While screening for early literacy risk has been an important element of other initiatives, a growing focus on the incidence of dyslexia in particular has led to a need for dyslexia screening in the early grades. As Hulme and Snowling (2016) note:

> "*Children with decoding difficulties/dyslexia experience deficits in phoneme awareness, letter-sound knowledge and rapid automatized naming in the preschool years and beyond. These phonological/language difficulties appear to be proximal causes of the problems in learning to decode print in dyslexia*" (p. 731).

The IDA has been influential in translations of dyslexia research into policy. The IDA notes in their definition that dyslexia is "…characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language*"* (IDA, n.d.). State educational agencies typically operationalize these characteristics as content requirements for dyslexia screening tools. These requirements typically include some or all of the following:

- Phonological/phonemic awareness
- Letter sound recognition
- Alphabet knowledge / letter naming
- Decoding skills / phonics / word level fluency
- Rapid naming
- Encoding skills
- Oral reading fluency

Research has long supported the value of early literacy screening. Early detection and subsequent intervention matter; together they can reduce the incidence of future reading failure (Snow et al., 1998; Torgesen, 2000). In particular, the IDA provides evidence that interventions marked by characteristics of *structured literacy* instruction are effective for students with dyslexia (IDA, 2015). Critical content elements include phonology, sound-symbol association, and syllable-level phonics. Critical elements of delivery include teaching that is systematic, explicit, and individualized.

7.2. Rapid Automatized Naming (RAN)

A RAN measure indicates the speed of correctly naming digits, letters, colors, or pictures of commonly known objects. Half a century ago, Denckla and colleagues found that a relevant kind of automaticity and speed could be gauged even before a student was decoding words: *rapid automatized naming (RAN)* of colors or objects (Denckla, 1972; Denckla & Rudel, 1976). In a 2015 meta-analysis, Araújo and colleagues confirmed that across studies, RAN scores have a moderate to strong correlation to reading outcomes, particularly reading fluency (Araújo et al., 2015). In particular, RAN predicts reading growth. Several studies have found that students with poor RAN performance tend to show lower growth in reading (Lervåg & Hulme, 2009; Al Otaiba & Fuchs, 2002). This tends to hold true even in the face of well-designed reading instruction: in Nelson and colleagues' meta-analytic review of reading intervention studies, RAN was the strongest predictor of treatment effectiveness, or student growth in response to intervention (Nelson et al., 2003).

For these reasons, the MAP Reading Fluency Dyslexia Screener produces a RAN score. While some have investigated *discrete*, one-by-one presentation of the item to be named, the literature is increasingly converging on a *serial* presentation (Georgiou et al., 2013). In *serial RAN*, several items are presented at once in rows. The task is to name the items left to right and line by line, in the same way that words are read on a page.

The items students are asked to name can be either *alphanumeric* (letters or numbers) or *non-alphanumeric* (colors or common objects). While alphanumeric measures have correlated more strongly to reading outcomes, young children who may not yet know all their letters or numbers have often been assessed using a non-alphanumeric RAN measure (Kirby et al., 2010). In MAP Reading Fluency, the RAN measure is *serial* and *non-alphanumeric*: across rows; students name simple common objects. For example, students will see a series of screens like the one in Figure 7.1 and will be asked to say the names of the pictures out loud, in left-to-right, top-to-bottom order.

Table 7.1. Specifications—Rapid Automatized Naming (RAN)

Code	029
Specifications	Five simple objects with one-syllable, common names were used: ball, car, hat, sock, tree. Two sets of 18 objects, appearing in random order minus any immediate repetitions, are presented for the student to name in order aloud. Three rows of six objects are on each screen, after the demonstration. Labels for each picture are presented and practiced before the RAN measure begins. Students' out-loud naming is scored automatically via speech scoring.
Item Pool	36 pictures to be named. Each picture is one of the five objects.
Duration	Student speed of naming is collected. Maximum duration on each screen is 45 seconds.

Figure 7.1. Sample Item—Rapid Automatized Naming (RAN)

Rapid Automatized Naming (RAN)	Name the objects aloud in order.	

It is important to note that RAN is different from other screening measures: while promising as a predictor, it should not be a target of instruction. In a 2010 review, Kirby et al. (2010) determined that there is insufficient evidence that naming speed is responsive to direct instruction. Further, they note that "[t]here is not yet a strong case for instruction to improve naming speed" (p. 356). Norton and Wolf (2012) echoed this summary: "[M]ost researchers would agree that training students on a RAN task would not be the optimal way to improve their reading fluency" (p. 446). For this reason, data on RAN are not attached to instructional "next steps" or instructional materials in MAP Reading Fluency.

7.3. Predictive Modeling

NWEA researchers created logistic regression models using spring *At-Risk* status on MAP Growth Reading as the outcome variable and Sentence Reading Fluency and the Foundational Skills domain scores as the predictor variables. Separate models were run by grade and term of the predictor variables (e.g., Spring Grade 2 with fall predictors). The *At-Risk* cut points for the spring MAP Growth Reading assessment were set at the 10th percentile of each grade for Grades K–3, as shown in Table 7.2. There were insufficient data from the RAN measure to include it in this round of modeling.

Table 7.2. MAP Growth Reading Spring Thresholds (External Criterion)

Grade	RIT Threshold	Percentile Rank
K	138	10
1	153	10
2	166	10
3	176	10

7.4. Classification Accuracy and AUC based on Model Predictions

The area under the ROC curve (AUC) was used to evaluate how much the classification model is capable of distinguishing between the at-risk and non-risk students. NCII recommends that the lower limit of the 95% confidence interval for the AUC be ≥ 0.80 for classification accuracy analyses (NCII, 2020). The lower limit of the confidence interval for these predictions was ≥ 0.80 for all grades and terms, except for Grade 3 in fall and winter and kindergarten in the winter, as shown in Table 7.3.

Table 7.3. AUC for Spring MAP Growth At-Risk Predictions

Term	Grade	AUC	LL	UL
Fall 2020-2022	K	0.81	0.80	0.82
	1	0.89	0.88	0.89
	2	0.87	0.86	0.88
	3	0.88	0.87	0.89
Winter 2020-2022	K	0.87	0.87	0.88
	1	0.90	0.90	0.91
	2	0.88	0.87	0.90
	3	0.89	0.88	0.90
Spring 2020-2022	K	0.90	0.89	0.90
	1	0.90	0.89	0.90
	2	0.90	0.89	0.91
	3	0.89	0.88	0.91

Table 7.4 presents the accuracy rate (i.e. correct classification rate) as well as sensitivity and specificity. The model was fine-tuned for each grade and term to achieve optimal classification accuracy, taking into account scores of all four predictors. Instead of setting cut points on individual domain scores and Sentence Reading Fluency scores, they were determined based on estimated at-risk probabilities. Cut points were selected to balance sensitivities and specificities. The classification results for most grades and terms achieved sensitivity and specificity statistics of ≥ 0.80, meeting the NCII rubrics for convincing evidence. For an explanation of sensitivity and specificity, see Section 6.4 of this technical report.

NWEA routinely conducts rigorous research to ensure high classification accuracy of the predictive model. The threshold on the criterion measure was updated to the 10th percentile of Spring MAP Growth outcome in Fall 2023 based on the latest research to improve identification of at-risk students. Consequently, the model currently predicts that approximately 25%-29% of students across grades and terms fall into the at-risk category. This update achieved significantly better classification results and reduced false positive rates compared to the previous model iteration. The continuous improvement of the predictive model provides educators with precise information to inform instructional and intervention decisions.

Table 7.4. Classification Results for the Predictive Model

Term	Grade	Proportion Correct	Sensitivity	Specificity	False Positive	False Negative	Proportion Flagged
Fall 2020-2022	K	0.75	0.74	0.75	0.25	0.27	0.29
	1	0.82	0.81	0.82	0.18	0.19	0.26
	2	0.79	0.78	0.80	0.21	0.22	0.26
	3	0.80	0.81	0.79	0.21	0.19	0.27
Winter 2020-2022	K	0.80	0.80	0.81	0.20	0.21	0.25
	1	0.83	0.81	0.83	0.17	0.19	0.29
	2	0.80	0.80	0.80	0.20	0.21	0.27
	3	0.80	0.83	0.79	0.21	0.17	0.26
Spring 2020-2022	K	0.82	0.83	0.82	0.19	0.18	0.25
	1	0.83	0.81	0.83	0.17	0.19	0.26
	2	0.80	0.84	0.80	0.20	0.16	0.26
	3	0.81	0.85	0.80	0.20	0.15	0.27

8. References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. MIT Press.

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special education*, *23*(5), 300–316.

Allington, R. L., McCuiston, K., & Billen, M. (2015). What research says about text complexity and learning to read. *The Reading Teacher*, *68*(7), 491–501.

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.

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. In D. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (vol. 2, pp. 41–51). Guilford.

Boerman-Cornell, W. (2016). The intersection of words and pictures: Second through fourth graders read graphic novels. *The Reading Teacher*, *70*(3), 327–335.

Cassady, J. C., & Smith, L. L. (2004). Acquisition of blending skills: Comparisons among body-coda, onset-rime, and phoneme blending tasks. *Reading Psychology*, *25*(4), 261–272.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, *49*(2), 278–293.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services In Schools, 32*, 38–50.

Clay, M. M. (1989). Concepts about print in English and other languages. *The Reading Teacher, 42*(4), 268–276.

Clemens, N. H., Shapiro, E. S., & Thoemmes, F. (2011). Improving the efficacy of first grade reading screening: An investigation of word identification fluency with other early literacy indicators. *School Psychology Quarterly*, *26*(3), 231.

Common Core State Standards Initiative. (2010). *Common core state standards for English Language arts & literacy in history/social studies, science, and technical subjects*. CCSSO & National Governors Association.

Cornelius, D. (2020). Interpreting pictures: Bridging the literacy gap with graphic novels. *Perpetua*, 19.

Cummings, K. D., Kaminski, R. A., Good, R. H., & O'Neil, M. (2011). Assessing phonemic awareness in preschool and kindergarten: Development and initial validation of first sound fluency. *Assessment for Effective Intervention*, *36*(2), 94–106.

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.

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading* (NCES 2006–469). U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Government Printing Office.

Deeney, T. A. (2010). One-minute fluency measures: Mixed messages in assessment and instruction. *The Reading Teacher*, *63*(6), 440–450.

Denckla, M. B. (1972). Color-naming defects in dyslexic boys. *Cortex*, *8*(2), 164–176.

Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471--479.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*(3), 219–232.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy (*pp. 3–40). Lawrence Erlbaum Associates Publishers.

Ehri, L. (2002). Phases of acquisition in learning to read words and implications for teaching. In R. Stainthorp & P. Tomlinson (Eds.), *Learning and teaching reading*. British Journal of Educational Psychology Monograph Series II.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, *18*(1), 5–21.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, *36*(3), 250–287.

Foorman, B. R., Fletcher, J. M., Frances, D. J., Carlson, C. D., Chen, D., & Mouzaki, A. (1998). *Technical report: Texas primary reading inventory*. Center for Academic and Reading Skills and the University of Houston.

Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in kindergarten through grade 2. *Reading and Writing*, *28*(5), 655–681.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301–330). Lawrence Erlbaum Associates.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*(2), 188–192.

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.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239–256.

García, J. R, & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84*(1), 74–111.

Georgiou, G. K., Parrila, R., Cui, Y., & Papadopoulos, T. C. (2013). Why is rapid automatized naming related to reading? *Journal of Experimental Child Psychology*, *115*(1), 218–225.

Gentry, J. R. (1982). An analysis of developmental spelling in “GNYS AT WRK.” *The Reading Teacher*, *36*(2), 192–200.

Gillon, G. T. (2004). *Phonological awareness: From research to practice*. Guilford Press.

Gillon, G. T. (2017). *Phonological awareness: From research to practice (second edition)*. Guilford Press.

Good, R. H., Kaminiski, R. A., Cummings, K., Dufour-Martel, C., Peterson, K., Powell-Smith, K., Stollar, S., & Wallin, J. (2011). *DIBELS/Next technical manual*. Dynamic Measurement Group.

Good, R. H., Kaminiski, R. A., Dewey, E., Wallin, J., Powell-Smoth, K. A., & Latimer, R. (2013-2019). *Acadience reading K-6 technical manual*. Dynamic Measurement Group.

Goswami, U., & Mead, F. (1992). Onset and rime awareness and analogies in reading. *Reading Research Quarterly*, 153–162.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *RASE: Remedial & Special Education, 7*, 6–10.

Guthrie, J. T., & Seifert, M. (1977). Letter–sound complexity in learning to identify words. *Journal of Educational Psychology, 69*(6), 686.

Han, K. T. (2016). Maximum likelihood score estimation method with fences for short-length tests and computerized adaptive tests. *Applied Psychological Measurement*, *40*(4), 289–301. https://doi.org/10.1177/0146621616631317

Hasbrouck, J. & Tindal, G. (2017). *An update to compiled ORF norms* (technical report no. 1702). Behavioral Research and Teaching, University of Oregon.

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly*, *15*(1), 75–90.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology*, *82*(1), 2–28.

Hulme, C., & Snowling, M. J. (2016). Reading disorders and dyslexia. *Current Opinion in Pediatrics*, *28*(6), 731–735.

International Dyslexia Association (IDA). (n.d.). *Definition of dyslexia*. https://dyslexiaida.org/definition-of-dyslexia/

International Dyslexia Association (IDA). (2015). *Effective reading instruction for students with dyslexia*.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review*, *36*(4), 582–600.

Jenkins, J. R., Fuchs, L. S., Van Den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of educational psychology*, *95*(4), 719.

Jiban, C., & Simpson, M. A. (2021). *MAP Reading Fluency theory of action*. NWEA. https://www.nwea.org/uploads/2021/06/MAP-Reading-Fluency-Theory-of-Action-Infographic-AUG21.pdf

Justice, L. M., & Ezell, H. K. (2002). Use of storybook reading to increase print awareness in at-risk children. *American Journal of Speech-Language Pathology, 11*(1), 17–29.

Justice, L. M., & Ezell, H. K. (2004). Print referencing: An emergent literacy enhancement strategy and its clinical applications. *Language, Speech, & Hearing Services in Schools, 35*(2), 185–193.

Katzir, T., Kennedy, B., Kim, Y., Lovett, M., Morris, R., & Wolf, M. (2006). The relationship of spelling recognition, RAN, and phonological awareness to reading skills in older poor readers and younger reading-matched controls. *Reading and Writing, 19*, 845–872.

Kilpatrick, D. A. (2012a). Phonological segmentation assessment is not enough: A comparison of three phonological awareness tests with first and second graders. *Canadian Journal of School Psychology, 27*(2), 150–165.

Kilpatrick, D. A. (2012b). Not all phonological awareness tests are created equal: Considering the practical validity of phonological manipulation versus segmentation. *Communique*, *40*(6), 31–33.

Kilpatrick, D. A. (2018). Incorporating recent advances in understanding word-reading skills into SLD diagnoses: The case of orthographic mapping. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment–fourth edition: Theories, tests, and issues*. Guilford.

Kim, Y. S., Wagner, R. K., & Foster, E. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific Studies of Reading*, *15*(4), 338–362.

Kirby, J. R., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly*, *45*(3), 341–362.

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, *100*(2), 310.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, linking and scaling: Methods and practices* (3d ed.). Springer.

Kroese, J. M., Hynd, G. E., Knight, D. F., Hiemenz, J. R., & Hall, J. (2000). Clinical appraisal of spelling ability and its relationship to phonemic awareness (blending, segmenting, elision, and reversal), phonological memory, and reading in reading disabled, ADHD, and normal children. *Reading & Writing: An Interdisciplinary Journal, 13*, 105–131.

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, *45*(2), 230–251.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323.

Lenchner, O., Gerber, M. M., & Routh, D. K. (1990). Phonological awareness tasks as predictors of decoding ability: Beyond segmentation. *Journal of Learning Disabilities, 23*(4), 240–247.

Lennon, C., & Burdick, H. (2014). *The Lexile Framework as an approach for reading measurement and success*. MetaMetrics, Inc. white paper. https://cdn.lexile.com/cms_page_media/135/The%20Lexile%20Framework%20for%20Reading.pdf

Leslie, L., & Caldwell, J. (2006). *Qualitative reading inventory – 4*. Allyn & Bacon.

Lepola, J., Lynch, J., Kiuru, N., Laakkonen, E., & Niemi, P. (2016). Early oral language comprehension, task orientation, and foundational reading skills as predictors of grade 3 reading comprehension. *Reading Research Quarterly*, *51*(4), 373–390.

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*(8), 1040–1048.

Liao, C. N., Chang, K. E., Huang, Y. C., & Sung, Y. T. (2020). Electronic storybook design, kindergartners' visual attention, and print awareness: An eye-tracking investigation. *Computers & Education*, *144*, 103703.

Linacre, J. M. (2021). Winsteps® Rasch measurement computer program (Version 4.8.0). Portland, Oregon: Winsteps.com

Lipson, M. Y., & Wixson, K. K. (2012). *Assessment and instruction of reading and writing difficulties: An interactive approach*. Pearson Higher Ed.

MetaMetrics. (2021). *Lexile® framework for oral reading: Development and validation*.

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322.

Mesmer, H. A. E., & Williams, T. O. (2015). Examining the role of syllable awareness in a model of concept of word: Findings from preschoolers. *Reading Research Quarterly*, *50*(4), 483–497.

Middaugh, D. (2019, April 17). *What we know about 2018 graphic novel sales*. Publisher's Weekly. https://www.publishersweekly.com/pw/by-topic/industry-news/comics/article/79818-what-we-know-about-2018-graphic-novel-sales.html.

Moats, L., & Tolman, C. (2009). The development of phonological skills. Excerpted from *Language essentials for teachers of reading and spelling (LETRS): The speech sounds of English: phonetics, phonology, and phoneme awareness* (module 2). Sopris West. http://www.readingrockets.org/article/development-phonological-skills.

Mokhtari, K., & Thompson, H. B. (2006). How problems of reading fluency and comprehension are related to difficulties in syntactic awareness among fifth graders. *Reading Research and Instruction, 46*, 73–94.

Mol, S. E., Bus, A. G., De Jong, M. T., & Smeets, D. J. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development*, *19*(1), 7–26.

Morris, D., Bloodgood, J. W., Lomax, R. G., & Perney, J. (2003). Developmental steps in learning to read: A longitudinal study in kindergarten and first grade. *Reading Research Quarterly*, *38*(3), 302–328.

National Center on Intensive Intervention (NCII). (2020). *Academic screening tools chart rating rubric.* American Institutes for Research.

National Governors Association Center for Best Practices & Council of Chief State School Officers (CCSSO). (2010). *Common core state standards*. http://www.corestandards.org/read-the-standards/

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. *Reports of the subgroups*. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf.

Nagy, W. E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52–77). Guilford Press.

Nelson, J., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice*, *18*(4), 255–267.

Newman, H. M. (2009). *The influence of mandated oral reading fluency assessments on teachers' thinking and practice*. Unpublished doctoral dissertation, University of Washington, Seattle.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*, 427–452.

Nunes, T., Bryant, P., & Barros, R. (2012). The development of word recognition and its significance for comprehension and fluency. *Journal of Educational Psychology*, *104*(4), 959.

NWEA. (2019). *MAP Reading Fluency technical report*.

NWEA. (2021a). *Spanish MAP Reading Fluency content guide*. NWEA Research Report. https://www.nwea.org/content/uploads/2022/01/Spanish-MAP-Reading-Fluency-Content-Guide-2021-08-04.pdf

NWEA. (2021b). *MAP Reading Fluency administration guidance document*. https://teach.mapnwea.org/impl/MRFAdministrationGuidance.pdf

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, *3*(2), 159–197.

O'Neill, T., Peabody, M., Tan, R., & Du, Y. (2013). How much item drift is too much? *Rasch Measurement Transactions, 27*(3), 1423–1424.

Paige, D., Rasinski, T., Magpuri-Lavell, T, & Smith, G. (2014). Interpreting the relationships among prosody, automaticity, accuracy, and silent reading comprehension in secondary students. *Journal of Literacy Research, 46*(2), 123–156.

Paulson, L. H. (2004). *The development of phonological awareness skills in preschool children: From syllables to phonemes*. Dissertation Abstracts International, University Microfilms No. 3166292. https://scholarworks.umt.edu/cgi/viewcontent.cgi?article=10558&context=etd.

Perfetti, C. A. (1997). The psycholinguistics of spelling and reading. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 21–38). Erlbaum.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal. *Merrill-Palmer Quarterly, 33*, 283–319.

Piasta, S. B., Justice, L. M., McGinty, A. S., & Kaderavek, J. N. (2012). Increasing young children's contact with print during shared reading: Longitudinal effects on literacy achievement. *Child Development*, *83*(3), 810–820.

Pirani-McGurl, C. A. (2009). *The use of item response theory in developing a Phonics Diagnostic Inventory*. University of Massachusetts Amherst.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*.

Rasinski, T. V., Reutzel, D. R., Chard, D., & Linan-Thompson, S. (2011). Reading fluency. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4). Routledge.

Read, C. (1971). Pre-school children's knowledge of English phonology. *Harvard Educational Review, 41*, 1–34.

Richgels, D. (1995). Invented spelling ability and printed word learning in kindergarten. *Reading Research Quarterly, 30*(1), 96–109.

Robbins, K. P., Hosp, J. L., Hosp, M. K., & Flynn, L. J. (2010). Assessing specific grapho-phonemic skills in elementary students. *Assessment for Effective Intervention*, *36*(1), 21–34.

Samejima, F. (1977). A use of the information function in tailored testing. *Applied Psychological Measurement, 1*(3), 233–247.

Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modifications. *Applied Psychological Measurement, 18*(3), 229–244.

Samuels, S. J. (2006). Toward a model of reading fluency. In S. J. Samuels & A. E. Farstrup (Eds), *What research has to say about fluency instruction* (pp. 24–46). International Reading Association.

Samuels, S. J. (2007). The DIBELS tests: Is speed of barking at print what we really mean by reading fluency? *Reading Research Quarterly, 42*(4), 563–566.

Scarborough, H. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. Shapiro, P. Accerdo, & A. Capute (Eds*.), Specific reading disability: A view of the spectrum* (pp. 75–119). York Press.

Schrank, F. A., Mather, N., & Woodcock, R. W. (2004). *Woodcock Johnson III Diagnostic Reading Battery: Comprehensive manual.* Riverside.

Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In D. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (vol. 2, pp. 173–182). Guilford Press.

Shanahan, T., Fisher, D., & Frey, N. (2016). The challenge of challenging text. *On Developing Readers: Readings from Educational Leadership (EL Essentials)*, 100.

Shanahan, T., & Lonigan, C. J. (2010). The National Early Literacy panel: A summary of the process and the report. *Educational Researcher*, *39*(4), 279–285.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.) .(1998). *Preventing reading difficulties in young children*. National Academy Press.

Speece, D. L. (2005). Hitting the moving target known as reading development some thoughts on screening children for secondary interventions. *Journal of Learning Disabilities*, *38*(6), 487–493.

Speece, D. L., & Case, L. P. (2001). Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology*, *93*(4), 735.

Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *The Journal of Special Education*, *36*(4), 223–233.

Stahl, S. A., & Heubach, K. M. (2005). Fluency-oriented reading instruction. *Journal of Literacy Research*, *37*(1), 25–60.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, *15*(1), 55–64.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*(2), 107.

Valencia, S. W., & Buly, M. R. (2004). Behind test scores: What struggling readers really need. *The Reading Teacher*, *57*(6), 520–531.

Valencia, S. W., Smith, A. T., Reece, A. M., Li, M., Wixson, K. K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly*, *45*(3), 270–291.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *Test of silent reading efficiency and comprehension*. Pro-Ed.

Walton, P. D., & Walton, L. M. (2002). Beginning reading by teaching in rime analogy: Effects on phonological skills, letter-sound knowledge, working memory, and word-reading strategies. *Scientific Studies of Reading*, *6*(1), 79–115.

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, *41*(2), 85–120.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA Press.

University of Oregon (2018-2020). *8th Edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS®): Technical manual*. University of Oregon. https://dibels.uoregon.edu

U.S. Census Bureau. (2016). *Geographic levels*. https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html#par_textimage_34

U.S. Department of Education, National Center for Education Statistics. (n.d.) *Common Core of Data (CCD): State nonfiscal public elementary/secondary education survey, 2020-21 v.1a* [Data set]. https://nces.ed.gov/ccd/elsi/

Xue, K. (2020). *PyTHETA* (version 1.0.0) [Computer software]. NWEA.

Appendix A: Descriptive Metadata for Oral Reading Passages

Passage Title	Lexile® Text Measure	Lexile® Oral Readability	Word Count
Jump Rope	180	20	168
Hello, Play!	180	160	188
The Class Pet	180	150	176
The Box	180	40	189
That's No Bug	190	-40	183
Little Cat	190	-50	178
We Can Win	190	180	178
Art Bin	190	-30	148
Mack the Cat	200	-70	157
Pink the Pig	200	70	187
Bus Stop	210	10	174
Art on a Plate	210	230	178
Ann's Bear	210	10	187
Sal Gets Wet	210	-10	167
He Plants Trees	220	120	185
Robot in School	220	270	191
Bird Nests	220	140	188
Zack in Rain	220	80	188
Paper Jet	250	220	194
Ben's Flag	270	50	171
Fancy Pie	270	330	190
Tell Time	290	140	197
Bats and Birds	330	330	208
Grandma's Cart	340	300	212
John Loved the Moon	350	220	204
Snack Time	360	200	199
Lost Coat	370	320	205
Sore Throat	370	260	212
Rainy Day	370	320	204
Lollipops	370	320	212
Alex's Collection	380	300	211
Be a Teacher	380	350	207
Swim the Channel	380	370	202
Crickets	390	390	204
The Class Garden	390	420	200
Skate	390	210	200
Toes That Show	390	300	206
Homes Around the World	390	440	205
My Teacher's Clothes	390	390	198
Sugar Maples	390	490	206
Losing Teeth	400	340	205
Bears	410	420	200
Zoo	430	310	204
Casey's Walk	440	290	208

Passage Title	Lexile® Text Measure	Lexile® Oral Readability	Word Count
Jay and Gus	450	290	208
Cleaning Our Room	460	220	201
Parker the Peacock	460	190	206
Bike Ride	470	260	201
Airplanes	470	330	206
Game Inventor	480	430	205
Bing the Polar Bear	480	350	207
Bus Love	490	310	201
Baker Brother	490	400	249
Old Photos	490	450	221
Pam and the Toy Chest	490	440	241
Drinking Fountain	490	480	237
Powwow	490	610	215
The Baseball	500	290	222
Playground Alien	500	420	220
Hamster on the Loose	500	460	212
Spell Pizza	510	370	236
Popcorn Science	520	450	215
Family Bowling	530	340	222
Blue Whales	530	440	251
Winter School Day	540	540	220
Class Trip	540	530	255
Training a Puppy	540	630	216
Lara the Inventor	550	660	222
Emperor Penguins	560	630	219
Bad Talent Show	560	640	221
Ants	570	660	212
Tree House	580	590	222
Butterflies and Moths	580	660	223
Dad Versus Socks	590	580	221
Music Museum	600	680	221
Kangaroos	600	770	217
The Family Blanket	600	680	217
Grandma Babysits	610	720	221
Grandpa and the Salt Mine	610	600	214
Mom's Performance	610	660	219
Truffles	610	730	220
Art in the Park	610	770	209
Field Mice	610	720	210
Movie Magic	620	640	222
Bird-watchers	620	780	209
Vacuum Cleaners	620	670	215
Vanilla	620	670	216
Monster Baby	630	700	219
Bricks	710	700	218
Rubber Bands	720	750	

Passage Title	Lexile® Text Measure	Lexile® Oral Readability	Word Count
Shopping for Food	740	700	213
Pizza History	740	780	220
Lava Monster	750	600	216
State Park	750	710	216
Owls	760	840	209
Dogs	760	720	207
What's In the Mirror	770	790	220
Waiting for Dad	810	600	231
The Surprise	820	850	229
Puffballs	840	880	231
Toby Comes Home	840	740	230
The Art of Juggling	860	780	224
Tumbleweeds	880	770	229
Mystery Bike Ride	900	800	232
Salt Mountain	900	990	227
Pink What?	920	770	212
The Kite	930	910	228
Gongs	940	1030	227
The Paint Vote	270	200	195
Party Clothes	280	170	198
Crown of Gems	280	170	202
How Tall Can It Grow?	290	130	198
A Night at the Fair	290	260	200
Lemon Pie	370	310	210
Frank the Fox	380	220	211
Fort Living Room	380	400	200
Nice, Cold Drink	380	240	211
Hills Day Parade	390	520	210
A New Puppy	450	320	201
Name That Truck	480	560	217
Fish	490	340	210
Horses	490	540	220
Tides	510	500	209
Alpine Slide	540	560	220
Sea Turtles	560	660	223
Frogs and Toads	580	570	214
Lazy River	590	620	217
Global Sandwich	610	780	218
Lang the Ladybug	610	590	221
Story of Roller Skating	610	670	223
Hot Air Balloons	640	740	228
A Special Visitor	650	780	230
Ice Cold	650	750	213
Video Games	670	760	228
Space Project	700	880	227
Escape Artists	710	950	231

Passage Title	Lexile® Text Measure	Lexile® Oral Readability	Word Count
Mural	710	810	230
Best Friends	710	730	228
A Summer Project	730	710	212
Skyscraper	730	740	230
Ice Sounds	760	710	226
Frog or Prince	770	720	220
Gold or Just Golden?	770	860	230
Hanging Around	770	830	230
Spoon, Fork, Spork	790	750	230
Fearsome Fungus	790	800	227
Little Wild Thing	800	880	229
An Interesting Day	800	810	229
Ostriches	800	750	225
A New Puzzle	810	910	228
Shrunk in the Night	810	820	232
Herdwick Sheep	820	880	229
The Together Garden	830	930	230
An Amazing Air Show	840	920	224
Whale or Shark?	840	740	231
Vanishing Act	840	710	228
Memories	850	820	130
No, It's Not a Bee	870	900	220
What's on Your Tongue?	870	780	229
Flying Lemurs	890	760	213
Tap Tap Tap	890	730	227
Welcoming Grandpa	900	820	230
Great Barrier Reef	900	960	230
The Imagination Game	900	1160	210
Bird Spy	900	1070	229
Mimic	900	1140	230
A New Friend	910	1230	230
Night of the Bats	910	780	227
Shane's Shadow Show	920	960	229
Tooling Around	920	980	228
The Old Car	920	940	226
Ethiopia	920	960	228
Traveling Library	930	1250	211
How Alarming	940	910	227
Freaky Weather	940	880	229
A Whole Lot of Something	940	870	228
Group World Records	980	900	234
Buses, Old School and New	980	890	230

Appendix B: Descriptive Statistics of Raw Scores—Foundational Skills

Table B.1. Descriptive Statistics of Raw Scores—Foundational Skills, Fall 2020

Measure	Code	Grade K			Grade 1			Grade 2		
		N	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Initial Sound Matching	001	31,309	2.96	2.31	20,180	3.24	2.44	6,516	3.32	2.53
Letter Knowledge	002	43,470	7.57	5.01	13,287	8.82	5.01	2,834	8.74	5.37
Letter-Sound Fluency	003	47,512	4.11	3.39	16,272	6.16	3.86	3,875	6.52	4.17
Listening Comprehension	004	75,593	9.52	3.39	91,243	11.60	2.94	60,003	12.61	2.55
Picture Vocabulary	005	75,608	11.08	3.45	91,250	12.56	2.87	60,004	13.14	2.54
Decoding: CVC	007	31,105	6.59	5.00	76,098	9.75	5.67	56,907	13.69	6.39
Sentence Reading Fluency	008	37,840	8.32	5.54	80,929	9.42	6.29	87,003	14.61	7.10
Counting Syllables	017	27,145	4.44	3.18	9,723	4.50	3.45	2,760	4.25	3.65
Onset-Rime Blending	018	54,093	5.20	3.37	32,480	6.87	3.89	11,496	7.22	4.01
Phoneme Blending	019	65,691	3.23	2.54	88,004	5.62	3.52	59,305	7.09	3.71
Phoneme Counting	020	30,364	2.56	2.33	37,012	3.61	2.65	16,967	4.24	2.99
Phoneme Addition/Deletion	021	22,034	4.57	2.51	62,123	5.21	2.70	51,302	5.93	3.06
Phoneme Substitution	022	20,683	3.54	2.12	59,177	4.06	2.27	49,645	4.58	2.51
Word Families: Initial Letter	023	39,925	3.40	2.25	27,780	4.06	2.52	7,960	4.85	2.95
Build Words: One Letter	024	70,633	5.57	3.85	89,341	10.11	4.84	59,786	13.56	5.29
Build Words: CVC	025	21,084	2.65	3.32	40,314	6.28	4.63	14,821	7.19	4.87
Build Words: Single Syllable	026	16,052	3.31	3.86	58,308	4.26	3.64	51,016	7.19	4.86
Decoding: Single Syllable	027	15,986	7.09	5.68	58,112	7.79	6.02	50,909	12.60	7.33
Rhyme Completion	030	32,298	3.08	2.04	10,865	3.24	2.15	3,127	3.34	2.30
Measure	**Code**	**Grade 3**			**Grade 4**			**Grade 5**		
		N	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Initial Sound Matching	001	1,289	3.29	2.59	67	3.58	2.54	28	3.96	2.71
Letter Knowledge	002	424	8.66	5.69	32	9.06	5.79	20	12.00	4.81
Letter-Sound Fluency	003	607	6.44	4.24	38	6.11	3.99	22	7.41	4.27
Listening Comprehension	004	12,764	12.65	2.75	798	13.29	2.28	321	13.32	2.54
Picture Vocabulary	005	12,764	13.03	2.82	798	13.70	2.24	321	13.83	2.37
Decoding: CVC	007	12,274	13.58	5.51	765	16.75	7.19	300	20.56	8.36
Sentence Reading Fluency	008	41,508	18.00	6.74	18395	20.65	6.68	12,882	22.61	7.01
Counting Syllables	017	648	3.94	3.23	24	5.83	4.28	18	3.78	2.80
Onset-Rime Blending	018	2,389	6.34	3.94	110	7.20	4.15	44	5.98	4.40
Phoneme Blending	019	12,586	6.85	3.61	785	8.28	3.74	311	9.11	4.22
Phoneme Counting	020	3,291	4.28	3.09	142	4.33	3.17	34	5.56	3.87
Phoneme Addition/Deletion	021	10,955	6.00	2.99	714	7.97	3.46	283	8.87	4.07
Phoneme Substitution	022	10,646	4.62	2.46	701	5.74	2.99	281	6.26	3.05
Word Families: Initial Letter	023	1,286	5.20	3.09	66	4.91	2.76	25	4.60	3.91
Build Words: One Letter	024	12,707	13.9	4.82	792	16.94	6.09	318	19.10	7.31
Build Words: CVC	025	2,396	7.64	5.22	106	7.25	5.38	33	9.36	6.28
Build Words: Single Syllable	026	11,286	7.68	4.45	721	11.55	6.13	292	13.87	7.66
Decoding: Single Syllable	027	11,280	12.70	6.07	721	17.32	7.66	292	19.68	8.19
Rhyme Completion	030	772	3.24	2.23	41	3.61	2.72	29	3.17	1.91

Table B.2. Descriptive Statistics of Raw Scores—Foundational Skills, Winter 2021

Measure	Code	Grade K			Grade 1			Grade 2		
		N	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Initial Sound Matching	001	11,229	3.11	2.39	3,769	3.23	2.55	23,187	3.11	2.37
Letter Knowledge	002	7,273	8.79	5.22	1,890	9.02	5.64	24,397	8.08	4.95
Letter-Sound Fluency	003	9,046	6.37	4.08	2,484	6.52	4.27	27,868	5.38	3.77
Listening Comprehension	004	73,579	12.14	2.83	33,898	12.48	2.79	70,133	10.59	3.26
Picture Vocabulary	005	73,580	12.72	2.90	33,898	12.77	2.93	70,134	11.68	3.29
Decoding: CVC	007	63,235	10.69	5.47	31,809	12.90	5.73	44,370	7.07	4.61
Sentence Reading Fluency	008	88,460	10.85	6.80	74,474	16.13	7.59	52,913	8.28	5.71
Counting Syllables	017	5,734	4.42	3.49	1,918	4.35	3.62	15,817	4.54	3.34
Onset-Rime Blending	018	18,662	6.82	3.92	6,672	6.88	3.97	38,324	6.06	3.68
Phoneme Blending	019	69,835	6.72	3.84	33,351	7.19	3.78	65,978	4.34	3.14
Phoneme Counting	020	23,178	3.88	2.82	9,118	4.15	2.95	31,416	3.11	2.48
Phoneme Addition/Deletion	021	54,975	5.89	2.94	28,539	6.16	3.07	33,384	5.00	2.61
Phoneme Substitution	022	53,217	4.68	2.53	27,674	4.87	2.59	31,484	3.83	2.21
Word Families: Initial Letter	023	16,820	4.45	2.70	4,884	4.99	3.02	34,207	3.68	2.28
Build Words: One Letter	024	70,571	11.71	4.85	33,762	13.57	5.02	68,275	7.49	4.33
Build Words: CVC	025	28,802	7.33	5.06	9,112	7.65	5.22	31,849	4.98	4.43
Build Words: Single Syllable	026	51,732	5.30	3.87	28,291	6.96	4.49	26,570	3.06	3.13
Decoding: Single Syllable	027	51,612	8.04	5.41	28,241	11.14	6.18	26,453	5.45	4.71
Rhyme Completion	030	6,318	3.11	2.13	2,180	3.26	2.31	17,844	3.06	2.06
Measure	**Code**	**Grade 3**			**Grade 4**			**Grade 5**		
		N	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Initial Sound Matching	001	793	3.20	2.52	50	4.24	3.60	35	7.71	5.02
Letter Knowledge	002	331	8.60	5.81	32	10.88	6.44	35	15.34	6.19
Letter-Sound Fluency	003	446	6.36	4.23	37	7.24	4.73	37	10.89	5.41
Listening Comprehension	004	8,784	12.69	2.83	475	13.46	2.10	293	13.69	1.77
Picture Vocabulary	005	8,784	12.91	2.88	475	13.73	2.21	293	13.94	2.08
Decoding: CVC	007	8,412	13.68	5.59	441	17.81	6.81	257	20.96	7.10
Sentence Reading Fluency	008	36,259	19.19	6.97	15859	21.65	7.40	11,001	23.24	7.49
Counting Syllables	017	445	4.25	3.52	14	4.07	2.34	8	4.88	3.48
Onset-Rime Blending	018	1,491	6.35	3.91	73	8.26	4.56	49	10.02	4.95
Phoneme Blending	019	8,651	7.04	3.58	447	8.91	3.90	266	9.45	3.73
Phoneme Counting	020	1,979	4.21	3.04	71	5.17	2.79	32	6.50	5.06
Phoneme Addition/Deletion	021	7,655	6.40	3.07	414	8.35	3.65	251	8.80	3.54
Phoneme Substitution	022	7,462	5.01	2.59	408	6.17	2.92	251	6.18	2.95
Word Families: Initial Letter	023	850	5.29	3.12	51	6.96	4.30	39	9.41	5.05
Build Words: One Letter	024	8,742	14.25	5.07	468	17.18	5.77	292	19.57	6.37
Build Words: CVC	025	1,636	7.83	5.35	64	9.06	6.55	23	6.09	4.52
Build Words: Single Syllable	026	7,782	8.05	4.70	414	11.67	6.06	249	12.92	5.91
Decoding: Single Syllable	027	7,770	12.74	6.18	413	17.54	7.23	250	19.80	7.57
Rhyme Completion	030	513	3.22	2.21	35	4.57	3.17	33	7.18	4.28

Table B.3. Descriptive Statistics of Raw Scores—Foundational Skills, Spring 2021

Measure	Code	Grade K			Grade 1			Grade 2		
		N	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Initial Sound Matching	001	23,735	3.08	2.37	9,646	3.12	2.39	3,628	3.12	2.43
Letter Knowledge	002	21,790	8.34	5.10	6,518	8.72	5.39	1,896	8.82	5.65
Letter-Sound Fluency	003	25,439	5.65	3.94	8,248	6.39	4.18	2,548	6.68	4.35
Listening Comprehension	004	95,279	11.19	3.12	78,467	12.40	2.70	31,956	12.46	2.88
Picture Vocabulary	005	95,285	12.12	3.12	78,468	12.80	2.85	31,958	12.61	3.07
Decoding: CVC	007	71,055	8.96	5.42	65,300	12.17	6.10	29,853	13.06	5.70
Sentence Reading Fluency	008	76,982	8.81	6.02	106,659	13.49	7.37	92,439	18.03	7.69
Counting Syllables	017	14,474	4.60	3.44	5,437	4.57	3.62	2,082	4.73	3.83
Onset-Rime Blending	018	38,934	6.50	3.87	16,655	6.92	4.00	6,599	6.80	3.94
Phoneme Blending	019	90,234	5.53	3.70	71,111	7.22	3.94	31,403	7.07	3.73
Phoneme Counting	020	37,458	3.42	2.64	20,764	4.07	2.96	8,631	4.19	2.98
Phoneme Addition/Deletion	021	57,971	5.45	2.81	58,047	6.15	3.03	26,570	6.14	3.02
Phoneme Substitution	022	55,455	4.27	2.42	56,366	4.90	2.64	25,782	4.84	2.58
Word Families: Initial Letter	023	34,337	3.98	2.51	14,629	4.69	2.87	4,882	5.22	3.09
Build Words: One Letter	024	92,534	9.34	4.85	71,870	12.70	5.22	31,792	13.58	5.09
Build Words: CVC	025	43,531	6.23	4.84	24,635	7.37	5.19	8,981	7.46	5.26
Build Words: Single Syllable	026	51,519	4.03	3.55	55,171	6.40	4.50	26,257	6.94	4.49
Decoding: Single Syllable	027	51,288	6.52	5.45	55,110	10.15	6.61	26,223	11.31	6.19
Rhyme Completion	030	16,547	3.11	2.18	6,011	3.10	2.19	2,348	3.28	2.37
Measure	**Code**	**Grade 3**			**Grade 4**			**Grade 5**		
		N	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Initial Sound Matching	001	660	3.21	2.58	45	4.40	3.10	18	3.44	1.95
Letter Knowledge	002	285	8.39	5.78	35	13.57	5.28	13	10.54	5.94
Letter-Sound Fluency	003	411	6.55	4.35	39	8.56	4.77	16	8.13	4.79
Listening Comprehension	004	7,370	12.61	2.94	400	13.15	2.36	203	13.39	2.30
Picture Vocabulary	005	7,370	12.74	3.12	400	13.29	2.53	203	13.59	2.45
Decoding: CVC	007	7,038	14.06	5.74	363	18.35	7.56	189	20.06	7.92
Sentence Reading Fluency	008	37,399	20.49	7.22	16,298	23.12	7.59	10,599	24.08	7.92
Counting Syllables	017	406	4.21	3.67	11	5.55	3.88	5	3.40	2.51
Onset-Rime Blending	018	1,261	6.56	3.96	68	8.06	4.28	29	8.72	4.26
Phoneme Blending	019	7,248	7.15	3.59	375	8.69	3.93	193	9.50	3.71
Phoneme Counting	020	1,680	4.32	3.15	71	4.66	3.36	32	4.72	3.65
Phoneme Addition/Deletion	021	6,362	6.48	3.12	345	8.29	3.90	180	9.69	4.20
Phoneme Substitution	022	6,188	4.98	2.62	342	6.11	3.20	175	6.85	3.36
Word Families: Initial Letter	023	785	5.42	3.21	56	6.13	4.03	27	6.67	3.36
Build Words: One Letter	024	7,326	14.49	5.19	399	17.34	6.77	199	19.20	7.02
Build Words: CVC	025	1,502	8.03	5.56	59	6.56	5.65	31	7.32	5.81
Build Words: Single Syllable	026	6,443	8.15	4.83	340	12.21	6.56	171	14.74	7.40
Decoding: Single Syllable	027	6,429	13.03	6.36	338	18.25	8.14	170	20.77	7.68
Rhyme Completion	030	464	3.07	2.18	30	4.53	2.92	12	4.17	2.41