

MAP® Reading Fluency[™] Technical Report March 2019



© 2019 NWEA.

NWEA, MAP, and Measures of Academic Progress are registered trademarks, and MAP Skills, MAP Growth, MAP Reading Fluency are trademarks, of NWEA in the U.S. and in other countries. All rights reserved. No part of this document may be modified or further distributed without written permission from NWEA.

The names of other companies and their products mentioned are the trademarks of their respective owners.

Suggested citation: NWEA. (2019). *MAP*® *Reading Fluency*[™] *technical report*. Portland, OR: Author.

Table of Contents

Chapter 1: Introduction	. 1
1.1. MAP Reading Fluency Overview	. 1
1.2. Background	. 2
1.2.1. Literature Review	. 2
1.2.2. Design Rationale	. 2
1.2.2.1. Gauging Improvements in Oral Reading	. 3
1.2.2.2. Informing Instruction for Students Who Cannot Read Passages	. 3
1.2.2.3. Gauging Student Readiness for Oral Reading from Passages	. 4
Chapter 2: Test Design	. 5
2.1. Phonological Awareness	. 6
2.1.1. Level 1: Rhymes & Syllables	. 7
2.1.2. Level 2: Initial Sounds	. 8
2.1.3. Level 3. Phoneme Biending & Segmenting	10
2.2. Phonics & Word Recognition	יבי 12
2.2.1 Lovel 1: Letters & Sounds	17
2.2.1. Level 1. Letters & Sounds	14
2.2.3. Level 3: CVC Words	17
2.2.4. Level 4: One-Syllable Words	19
2.3. Sentence Reading Fluency	21
2.4. Language Comprehension	22
2.4.1. Picture Vocabulary	23
2.4.2. Listening Comprehension	23
2.5. Oral Reading	24
2.5.1. Oral Reading: Picture Book	24
2.5.2. Oral Reading: Passages and Comprehension Quiz	26
2.5.2.1. Passage Specifications	27
2.5.2.2. Passage Comprehension Quiz	28
Chapter 3: Item Development	29
3.1. Item Template Creation and Review	29
3.2. Item Writing and Review for Individual Items and Sets	29
Chapter 4: Test Administration	31
4.1. Administration Setup	31
4.2. Managing Students and Test Sessions	32
4.3. Pausing, Resuming, and Discarding In-progress Tests	32
4.4. Adaptive Presentation of Measures	32
Chapter 5: Scoring & Reporting	34
5.1. Score Classification and Performance Levels	34
5.1.1. Raw Score Conversion to Performance Levels: Oral Reading Measures	34
5.1.2. Instructional Reading Levels	35
5.1.3. Raw Score Conversion to Performance Levels: Foundational Skills Measures	36
5.2. Individual Student Reports	38

5.2.1. Reader Profiles and Foundational Skills Profiles	39
5.2.1.1. Oral Reading Rate	40
5.2.1.2. Decoding Accuracy	40
5.2.1.3. Passage Comprehension	41
5.3. LanguaMetrics' Speech Scoring Technology	41
Chapter 6: Technical Characteristics	42
6.1. Data Collection	42
6.2. Item Difficulty and Discrimination	44
6.3. Summary Raw Score Statistics	46
6.4. Reliability Evidence	49
6.4.1. Overview	49
6.4.2. Summary Marginal Reliability Statistics	49
6.5. Concurrent Validity Evidence	50
6.6. Relationship between Sentence Reading Fluency and Oral Reading Fluency	54
6.7. Effectiveness of Sentence Reading Fluency in Classifying Oral Reading Fluency	54
6.8. Examining the Reliability of the LanguaMetrics Machine Scores	57
6.8.1. Handscoring Process	57
6.8.1.1. Rangefinding	57
6.8.1.2. Scorer Training	58
6.8.2. Study Results	59
References	61

List of Tables

Table 2.1. Assessed Measures in Each Domain	. 5
Table 2.2. Phonological Awareness Progression	. 6
Table 2.3. Specifications—Rhyming Words	. 7
Table 2.4. Specifications—Counting Syllables	. 8
Table 2.5. Specifications—Onset-Rime Blending	. 9
Table 2.6. Specifications—Initial Sound Matching	. 9
Table 2.7. Specifications—Phoneme Blending	10
Table 2.8. Specifications— Phoneme Counting	11
Table 2.9. Specifications—Phoneme Addition/Deletion	12
Table 2.10. Specifications—Phoneme Substitution	13
Table 2.11. Phonics & Word Recognition Progression	14
Table 2.12. Specifications—Letter Knowledge	14
Table 2.13. Specifications—Letter-Sound Fluency	15
Table 2.14. Specifications—Build Words: One Letter	16
Table 2.15. Specifications—Word Families: Initial Letter	17
Table 2.16. Specifications—Decoding: CVC	18
Table 2.17. Specifications—Build Words: CVC	18
Table 2.18. Specifications—Decoding: Single-Syllable	19
Table 2.19. Specifications—Build Words: Single Syllable	20

Table 2.20. Specifications—Sentence Reading Fluency	21
Table 2.21. Specifications—Picture Vocabulary	23
Table 2.22. Specifications—Listening Comprehension	24
Table 2.23. Specifications—Oral Reading: Picture Book	25
Table 2.24. Readability Measures and Word Count for Picture Book Texts	26
Table 2.25. Descriptive Metadata for Oral Reading Passages	28
Table 4.1. Stages Presented Within the Adaptive Oral Reading Test Format	33
Table 4.2. Measures and Reported Outcomes	33
Table 5.1. Expectation Levels for Oral Reading Fluency Based on WCPM On Grade-level Tex	t
	34
Table 5.2. Performance Levels based on Fall, Winter, and Spring WCPM	34
Table 5.3. Expectations for Instructional Reading Levels by Text Lexile	35
Table 5.4. Performance Levels for Decoding Accuracy Based on Percent Accuracy	36
Table 5.5. ZPD Levels for Phonological Awareness	37
Table 5.6. ZPD Levels for Phonics & Word Recognition	37
Table 5.7. Performance Expectations by ZPD Level	37
Table 5.8. Performance Expectations for Language Comprehension	38
Table 6.1. Sample Population Demographics (Winter 2017)	43
Table 6.2. Sample Population Demographics (Fall 2017)	43
Table 6.3. Sample Population Demographics (Winter 2018)	43
Table 6.4. Sample Population Demographics (Spring 2018)	44
Table 6.5. Mean <i>P</i> -Values (p_m) and Point-Biserial Correlation Coefficients (r_{pbm}) (Fall 2017)	45
Table 6.6. Mean <i>P</i> -Values (<i>p_m</i>) and Point-Biserial Correlation Coefficients (<i>r_{pbm}</i>) (Winter 2018)	45
Table 6.7. Mean <i>P</i> -Values (<i>p_m</i>) and Point-Biserial Correlation Coefficients (<i>r_{pbm}</i>) (Spring 2018)	46
Table 6.8. Summary Raw Score Statistics (Fall 2017)	47
Table 6.9. Summary Raw Score Statistics (Winter 2018)	47
Table 6.10. Summary Raw Score Statistics (Spring 2018)	48
Table 6.11. Summary Marginal Reliabilities for Foundational Skill Forms (Fall 2017, Winter	
2018, and Spring 2018)	50
Table 6.12. Pearson Product-Moment Correlation Coefficients (r) between MAP Reading	
Fluency Passage WCPM and Comprehension Scores vs. MAP Growth Scores (Fal	
2017)	51
Table 6.13. Pearson Product-Moment Correlation Coefficients (<i>r</i>) between MAP Reading	
Fluency Passage WCPM and Comprehension Scores vs. MAP Growth Scores (Winter 2018)	52
Table 6.14 Pearson Product-Moment Correlation Coefficients (A between MAP Reading	52
Fluency Passage WCPM and Comprehension Scores vs. MAP Growth Scores	
(Spring 2018)	53
Table 6.15. Pearson Product-Moment Correlation Coefficients (r) between Sentence Reading	
Fluency and Passage Oral Reading Fluency (Winter 2017)	54
Table 6.16. Odds Ratio Estimated from Logistic Regression across Passages and Grades	
(Winter 2017)	56
Table 6.17. Classification Accuracy Statistics: Using Raw Score of Sentence Reading Fluency	/
to Screen Passage Reader (Passage WCPM Cutoff Value = 30) (Winter 2017)	56

Table 6.18. IRR and IRA Score Reliability Re	sults: Machine vs. Human Scores for Passage
Reading Responses (Winter 2017	

List of Figures

Figure 2.1. Sample Item—Rhyming Words	7
Figure 2.2. Sample Item—Counting Syllables	8
Figure 2.3. Sample Item—Onset-Rime Blending	9
Figure 2.4. Sample Item—Initial Sound Matching	10
Figure 2.5. Sample Item—Phoneme Blending	11
Figure 2.6. Sample Item—Phoneme Counting	11
Figure 2.7. Sample Item—Phoneme Addition/Deletion	12
Figure 2.8. Sample Item—Phoneme Substitution	13
Figure 2.9. Sample Item—Letter Knowledge	15
Figure 2.10. Sample Item—Letter-Sound Fluency	15
Figure 2.11. Sample Item—Build Words: One Letter	16
Figure 2.12. Sample Item - Word Families: Initial Letter	17
Figure 2.13. Sample Item—Decoding: CVC	18
Figure 2.14. Sample Item – Build Words: CVC	19
Figure 2.15. Sample Item—Decoding: Single-Syllable	20
Figure 2.16. Sample Item—Build Words: Single Syllable	21
Figure 2.17. Sample Item—Sentence Reading Fluency	22
Figure 2.18. Sample Item: Picture Vocabulary & Listening Comprehension	22
Figure 2.19. Sample Item—Oral Reading: Picture Book	26
Figure 2.20. Sample Passage	27
Figure 2.21. Sample Item—Oral Reading: Passage Comprehension Quiz	27
Figure 5.1. Sample Student Report—Oral Reading Path	38
Figure 5.2. Sample Student Report—Foundational Skills Path	39
Figure 5.3. Reader Profile	40
Figure 5.4. Foundational Skills Profile	40
Figure 6.1. Classification of Oral Reading Fluency by Sentence Reading Fluency	55

List of Abbreviations

Below is a list of abbreviations that appear in this technical report.

BR	base rate
CCSS	Common Core State Standards
CVC	consonant-vowel-consonant
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
DRA	Developmental Reading Assessment
ELL	English language learner
FN	false negative
FNR	false negative rate
FP	false positive
FPR	false positive rate
HMM	hidden Markov model
ICC	intra-class correlation
IRA	inter-rater agreement
IRR	inter-rater reliability
MAP	Measures of Academic Progress®
OCR	overall classification rate
ROC	receiver operating characteristic
SEM	standard error of measurement
SEN	sensitivity
SEP	specificity
SME	Strategic Measurement and Evaluation, Inc.
TN	true negative
TP	true positive
WCPM	words correct per minute

Chapter 1: Introduction

This technical report documents the processes and procedures employed by NWEA® to build and support MAP® Reading Fluency[™], an oral reading fluency assessment that adapts to accommodate pre-readers, early readers, and independent readers in Grades K–3. MAP Reading Fluency is the first measure of oral reading fluency for beginning readers to use speech recognition-based scoring rather than relying on human judgment and individual administration. NWEA began piloting the MAP Reading Fluency assessment at several schools during 2016–2017, with more than 2,000 students tested. In the 2017–2018 school year, NWEA released MAP Reading Fluency in an early adopter program during which approximately 24,000 students participated. In 2018–2019, MAP Reading Fluency became available for purchase by the general public.

1.1. MAP Reading Fluency Overview

MAP Reading Fluency measures oral reading fluency (i.e., the ability to read text aloud quickly, accurately, and with good inflection); decoding accuracy (i.e., the ability to translate a printed letter or word into a sound); and literal comprehension (i.e., the ability to understand the meaning of a passage). These areas are evaluated based on oral reading of up to three passages, approximately 200 words each. From these results, a reader profile and recommended next steps are generated.

MAP Reading Fluency can be administered in a group setting rather than one-on-one. Students wear headsets with microphones and read the test content out loud. The audio is recorded, scored, and saved for future playback. To start, a narrator greets the students and confirms that they understand the directions. Each student reads a short picture book to get started. Then they 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. After reading, students answer selected-response items to demonstrate their comprehension. If the student is not ready to read passages, a series of measures are presented that assess foundational reading skills, including phonological awareness, early phonics and word recognition skills, listening comprehension, and picture vocabulary.

The test takes about 20 minutes and can be taken three times a year in the fall, winter, and spring. MAP Reading Fluency is automatically scored, with results appearing in the educator reporting site. For each student, the test provides the following:

- Oral reading fluency or foundational skills proficiency relative to grade-level expectations
- Individualized literacy profile
- Recommended next steps

For students who read the passages, the report shows the words correct per minute (WCPM), decoding accuracy, and reading 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 reported as Meeting, Exceeding, Approaching, or Below Expectation. For students who were not ready for passages, the foundational skills report shows their proficiency in decoding side skills and in oral language comprehension. Proficiencies in Phonological Awareness and Phonics and 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 Meeting, Exceeding, Approaching, or Below Expectation.

1.2. Background

1.2.1. Literature Review

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 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, Wallace, Wiley, Tichá, & Espin, 2007; Jenkins, Hudson, & Johnson, 2007). Particularly among students still building their reading comprehension skills, changes in oral reading fluency offer a valuable indicator of overall growth in reading proficiency (Fuchs, Fuchs, Hosp, & Jenkins, 2001). However, researchers have 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).

Early warnings about possible instructional implications of assessing WCPM without comprehension (e.g., Deno, 1985) began with a shift among some educators toward equating faster oral reading with better reading (Newman, 2009; Deeney, 2010). Many researchers currently assert that the construct of oral reading fluency includes prosody (i.e., a student's phrasing and expression in support of meaning) (Kuhn, Schwanenflugel, & Meisinger, 2010; Rasinski, Reutzel, Chard, & Linan-Thompson, 2011; Samuels, 2006). In this case, faster reading can even be at odds with better, more prosodic reading (Daane, Campbell, Grigg, Goodman, & Oranje, 2005).

Still, reading that has sufficient rate, accuracy, and prosody is not the end goal. The real goal is improving comprehension. Comprehension of text is harder where 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.

Echoing the model of informal reading inventories, some critics of one-minute WCPM measures argue that a more robust approach to assessing oral reading fluency allows students to read aloud a complete passage 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 reading comprehension: rate, accuracy, prosody, and comprehension. These four data points, they argue, are also those that best enable individualizing or differentiating instruction. This more robust approach is the model for the MAP Reading Fluency assessment of oral reading fluency.

1.2.2. Design Rationale

MAP Reading Fluency is designed to point oral reading fluency data at immediate instructional decisions. These include finding appropriate instructional emphases and levels of text for individual students and instructional groupings. WCPM, accuracy, and low-level comprehension are 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.

In contrast to typical oral reading fluency measures (e.g., DRA and DIBELS), MAP Reading Fluency combines data on a student's decoding accuracy and comprehension with their oral reading rate to generate a profile of strengths and needs in oral passage reading. Some students read at a fast rate but with poor accuracy on word decoding. 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.

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 WCPM score falls below the 25th percentile on published national norms (Hasbrouck & Tindal, 2017), a student's performance is flagged for deeper follow up.

The MAP Reading Fluency assessment accomplishes the following:

- Gauge improvement in oral reading
- Inform instruction for students who cannot read passages
- Gauge student readiness for oral reading from passages

1.2.2.1. 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 feels 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., when they show 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 then challenged to read from passages at a higher level of text complexity.

Within one level of text complexity, improvements in reading are shown by increases in WCPM, decoding accuracy, and basic comprehension of the passage. Simultaneously, improvements are shown by increases in the text complexity of the passages a student reads.

1.2.2.2. Informing Instruction for Students Who Cannot 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
- 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. Each of the skills assessed aligns to the Common Core State Standards (CCSS); within the Foundational Skills standards, MAP Reading Fluency assesses letter identification, all phonological awareness standards, and all phonics standards through Grade 2. Each step in the phonological awareness and phonics progressions is mapped to best practice instructional materials made available by the Florida Center for Reading Research.

1.2.2.3. 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, Mather, & Woodcock, 2004) and the Test of Silent Reading Efficiency and Comprehension (Wagner, Torgesen, Rashotte, & Pearson, 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, Wagner, & Foster, 2011).

Within MAP Reading Fluency, the classification accuracy of silent sentence reading in predicting oral reading fluency was evaluated empirically. Findings are reported in this report in Chapter 6: Technical Characteristics, Table 6.17.

Chapter 2: Test Design

MAP Reading Fluency consists of the following two test formats, each with several alternate forms that include the same operational item pools and different field test content. Table 2.1 presents the domains and measures included in each format. The measures presented in the table are described in detail in the following sections.

- 1. The **Adaptive Oral Reading** format contains all the measures in Table 2.1. Content is presented according to adaptive test logic based on student performance within the test session.
- 2. The Foundational Skills format includes all the measures in Table 2.1 except the Oral Reading measures. Students taking the Foundational Skills test only proceed to Sentence Reading Fluency if they demonstrate an ability to read individual decodable words. Each test event includes a subset of measures in Phonological Awareness and Phonics & Word Recognition, selected adaptively based on performance within a progression of skills.

The MAP Reading Fluency test design is based on the Simple View of Reading model, which proposes that two broad factors enable or limit comprehension: decoding and language comprehension (Gough & Tunmer, 1986). When decoding is weak, even a student with excellent oral language comprehension cannot fully comprehend the text. MAP Reading Fluency includes a set of measures focusing on knowledge and skills with print or sounds and the process of mapping print to sound (i.e., decoding). Measures in this domain range from letter knowledge and phonemic awareness to word and sentence level reading. For students at a stage where they are not ready for reading full passages, MAP Reading Fluency administers an adaptively selected subset of these foundational print, sound, and print/sound decoding measures instead.

					Assesse	ed By:
Domain	Measure	Code	Adaptive Oral Reading	Foundational Skills		
	Rhyming Words	015	✓	\checkmark		
	Counting Syllables	017	\checkmark	\checkmark		
_	Onset -Rime Blending	018	\checkmark	\checkmark		
Decoding:	Initial Sound Matching	001	\checkmark	\checkmark		
Awareness	Phoneme Blending	019	\checkmark	\checkmark		
, maionooo	Phoneme Counting	020	\checkmark	\checkmark		
	Phoneme Addition/Deletion	021	\checkmark	\checkmark		
	Phoneme Substitution	022	\checkmark	\checkmark		
	Letter Knowledge	002	\checkmark	\checkmark		
	Letter-Sound Fluency	003	\checkmark	\checkmark		
	Build Words: One Letter	024	\checkmark	\checkmark		
Decoding:	Word Families: Initial Letter	023	\checkmark	\checkmark		
Phonics & Word	Decoding: CVC	007	\checkmark	\checkmark		
Recognition	Build Words: CVC	025	\checkmark	\checkmark		
	Decoding: Single Syllable	027	\checkmark	\checkmark		
	Build Words: Single Syllable	026	\checkmark	\checkmark		
	Sentence Reading Fluency	008	\checkmark	\checkmark		

Table 2.1. Assessed Measures in Each Domain

			Assesse	ed By:
Domain	Measure	Code	Adaptive Oral Reading	Foundational Skills
Language	Picture Vocabulary	005	✓	\checkmark
Comprehension	Listening Comprehension	004	\checkmark	\checkmark
	Oral Sight Word Reading	006	✓	
Oral Deadinest	Oral Reading: Picture Book	013	\checkmark	
Oral Reading	Oral Reading: Passages	011	\checkmark	
	Oral Reading: Passage Comprehension Quiz	014	\checkmark	

*Oral Sight Word Reading was not included in 2018–2019. Oral Reading: Passages and Oral Reading: Passage Comprehension Quiz are administered as a set (e.g., students read a passage then answer items about it) and are collectively known as Passage Comprehension.

2.1. Phonological Awareness

Early learners' phonemic awareness is among the strongest predictors of future decoding proficiency in English (Gillon, 2004; Melby-Lervåg, Lyster, & Hulme, 2012). The skills children use in working with larger sounds 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. 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). However, it is useful to find children *earlier* who are not on track toward that ability. 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 MAP Reading Fluency measures of phonological awareness are designed to fit to this research-based progression, with two measures at each of four levels, as shown in Table 2.2.

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

Table 2.2. Phonological Awareness Progression

2.1.1. Level 1: Rhymes & 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 work with onset-rime blending and segmentation. Rhyme sensitivity strongly predicts later development of phonemic awareness skills (Anthony & Lonigan, 2005).

Table 2.3 presents the specifications for the Rhyming Words measure, and Figure 2.1 presents a sample Rhyming Words item. Table 2.4 presents the specifications for Counting Syllables, and Figure 2.2 presents a sample Counting Syllables item.

Code	015	
Specifications	In this speeded measure, students choose the two words that rhyme. 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 were rejected. Score is correct pair selections over 2 minutes.	
Item Pool	Up to 30 items presented in random order	
Duration	2 minutes, speeded	
CCSS Alignment	t K.RF.2.a – Recognize and produce rhyming words.	

Table 2.3. Specifications—Rhyming Words

Figure 2.1. Sample Item—Rhyming Words

Rhyming Words	Listen to four words said aloud. Choose two that rhyme.	

Table 2.4. Specifications—Counting Syllables

Code	017
Specifications	In this speeded measure, 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 20 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



2.1.2. 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, Kaminski, Good, & O'Neil, 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 2.5 presents the specifications for the Onset-Rime Blending measure, and Figure 2.3 presents a sample Onset-Rime Blending item. Table 2.6 presents the specifications for Initial Sound Matching, and Figure 2.4 presents a sample Initial Sound Matching item.

Table 2.5. Specifications—Onset-Rime Blending

Code	018
Specifications	In this speeded measure, 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 29 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

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

Table 2.6. Specifications—Initial Sound Matching

Code	001
Specifications	In this speeded measure, students select the two words with the same initial sound. Replayable 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 were not clearly depictable by a simple illustration were rejected. Score is correct pair selections over 2 minutes.
Item Pool	Up to 16 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



2.1.3. Level 3: Phoneme Blending & 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, Beck, Bell, & Hughes, 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 2.7 presents the specifications for the Phoneme Blending measure, and Figure 2.5 presents a sample Phoneme Blending item. Table 2.8 presents the specifications for Phoneme Counting, and Figure 2.6 presents a sample Phoneme Counting item.

Code	019
Specifications	In this speeded measure, 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 30 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—Phoneme Blending

	Listen to three separated			
Phoneme Blending	phonemes. Blend the sounds together and choose the word.			
		_=	-	

Table 2.8. Specifications— Phoneme Counting

Code	020
Specifications	In this speeded measure, 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 30 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.			2 Arrow	>		٦
	phonemes, count them and choose the	1	2	3	4	5	
	number.				C)))		
			-		A		

2.1.4. 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, Fey, Zhang, & Tomblin, 2001; Kilpatrick, 2012a; Kroese, Hynd, Knight, Hiemenz, & Hal, 2000; Lenchner, Gerber, & Routh, 1990). Phonemic skills at this level are developed, reciprocally, by practice with decoding words (Shanahan & Lonigan, 2010).

Table 2.9 presents the specifications for the Phoneme Addition/Deletion measure, and Figure 2.7 presents a sample Phoneme Addition/Deletion item. Table 2.10 presents the specifications for Phoneme Substitution, and Figure 2.8 presents a sample Phoneme Substitution item.

Table 2.9. Specifications—Phoneme Addition/Deletion

Code	021
Specifications	In this speeded measure, 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 I- 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 2.10. Specifications—Phoneme Substitution

Code	022
Specifications	In this speeded measure, 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 I- 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



2.2. 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 MAP Reading Fluency measures of phonics and word recognition 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 2.11.

Level 1: Letters & 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

 Table 2.11. Phonics & Word Recognition Progression

2.2.1. Level 1: Letters & 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, Mills, Ritchey, & Hillman, 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 2.12 presents the specifications for the Letter Knowledge measure, and Figure 2.9 presents a sample Letter Knowledge item. Table 2.13 presents the specifications for Letter-Sound Fluency, and Figure 2.10 presents a sample Letter-Sound Fluency item.

Table 2.12.	Specifications—L	etter Knowledge
-------------	------------------	-----------------

Code	002
Specifications	In this speeded measure, 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



Table 2.13. Specifications—Letter-Sound Fluency

Code	003
Specifications	In this speeded measure, 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.		o k	r a	U S	m p	1
			_=		-		

2.2.2. 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, Bloodgood, Lomax, & Perney, 2003). In English, words with the same rime (sometimes called "word families") offer an analogy-based route to early whole word decoding (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995; Walton & Walton, 2002).

Table 2.14 presents the specifications for the Build Words: One Letter measure, and Figure 2.11 presents a sample Build Words: One Letter item. Table 2.15 presents the specifications for Word Families: Initial Letter, and Figure 2.12 presents a sample Word Families: Initial Letter item.

Table 2.14. Specifications—Build Words: One Letter

Code	024
Specifications	In this speeded measure, 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 to30 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	K.RF.3 – Know and apply grade-level phonics and word analysis skills in decoding words. 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



Table 2.15. Specifications—Word Families: Initial Letter

Code	023
Specifications	In this speeded measure, 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 20 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. Identify the correct decoding of the unpaired word.	pig wig vig

2.2.3. 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, Fuchs, and Compton (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, Burns and Griffin (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, Hosp, Hosp, & Flynn, 2010; Nunes, Bryant, & Barros, 2012).

Table 2.16 presents the specifications for the Decoding: CVC measure, and Figure 2.13 presents a sample Decoding: CVC item. Table 2.17 presents the specifications for Build Words: CVC, and Figure 2.14 presents a sample Build Words: CVC item.

Table 2.16. Specifications—Decoding: CVC

Code	007
Specifications	In this speeded, 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 36 items presented in random order
Duration	1 minute, speeded
CCSS Alianment	1.RF.3.b – Decode regularly spelled one-syllable words.

Figure 2.13. Sample Item—Decoding: CVC



Table 2.17. Specifications—Build Words: CVC

Code	025
Specifications	In this speeded measure, 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



2.2.4. 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 graphophonemic patterns have been used in timed word reading fluency measures. Compared with other brief screening measures designed to flag Grade 1 students at risk of poor reading outcomes, word identification fluency is among the strongest (Clemens, Shapiro, & Thoemmes, 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 reading comprehension, even after the effects of word reading proficiency had been included.

Table 2.18 presents the specifications for the Decoding: Single-Syllable measure, and Figure 2.15 presents a sample Decoding: Single-Syllable item. Table 2.19 presents the specifications for Build Words: Single Syllable, and Figure 2.16 presents a sample Build Words: Single Syllable item.

Code	027
Specifications	In this speeded, 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 letters 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 is required to 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.

Table 2.18. Specifications—Decoding: Single-Syllable

Item Pool	Up to 30 items presented in random order
Duration	1 minute, speeded
CCSS Alignment	1.RF.3.bDecode 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	Choose the picture that matches the onscreen word.	leaf

Table 2.19. Specifications—Build Words: Single Syllable

Code	026
Specifications	In this speeded measure, 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; 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.bDecode 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



2.3. 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 (e.g., 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, sentence reading fluency measures are presented to all children to help discern possible readiness for oral passage reading.

Table 2.20 presents the specifications for the Sentence Reading Fluency measure, and Figure 2.17 presents a sample Sentence Reading Fluency item.

Table 2.20. S	Specifications—	-Sentence	Reading	Fluency
---------------	-----------------	-----------	---------	---------

Code	008
Specifications	In this speeded measure, 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 Grade 1 with word count ranging from 3 to 7 words. Score is correct selections over 2 minutes.
Item Pool	Up to 45 items presented in random order
Duration	2 minutes





2.4. Language Comprehension

In Gough and Tunmer's Simple View of Reading (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 reading comprehension as students mature (Foorman, Herrera, Petscher, Mitchell, & Truckenmiller, 2015; Lepola, Lynch, Kiuru, Laakkonen, & Niemi, 2016). While it is possible to assess reading 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 not reading passages orally
- 2. Students showing poor literal comprehension on lowest level passages

Figure 2.18 presents a sample item for the two language comprehension measures: Picture Vocabulary and Listening Comprehension.



Figure 2.18. Sample Item: Picture Vocabulary & Listening Comprehension

2.4.1. Picture Vocabulary

One aspect of a student's language comprehension is vocabulary knowledge. When a student produces the word in response to a picture, as in assessments such as the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007), expressive vocabulary is assessed. In MAP Reading Fluency, the focus is instead 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-K through Grade 1 strongly predicts reading comprehension by Grade 4 (Sénéchal, Ouelette, & Rodney, 2006; Scarborough, 1998; Cunningham & Stanovich, 1997). Table 2.21 presents the specifications for the Picture Vocabulary measure.

Code	005
Specifications	In this untimed measure, 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 Grade 1 vocabulary. Those not easily depicted in a simple illustration were rejected. On a culled list, feedback was elicited in two cycles from educators with kindergarten and Grade 1 expertise and English language learner (ELL) expertise. Words with meanings that varied culturally or with confusing cognates in Spanish were removed. Numerically equal word lists were established for Grade K and Grade 1 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 35
Duration	Untimed

2.4.2. Listening Comprehension

As noted in the 2008 National Early Literacy Panel report, 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 reading comprehension (Catts, Adlof, & Weismer, 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 reading comprehension, for Grades K–2. Table 2.22 presents the specifications for the Listening Comprehension measure.

Table 2.22. Specifications—Listening Comprehension

Code	004
Specifications	In this untimed measure, 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 Grade 1 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 Grade 1 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/Grade 1 expertise and ELL expertise.
Item Pool	15 items presented in randomly form a pool of 36
Duration	Untimed

2.5. Oral Reading

When students begin to read from connected text, a key focus for both instruction and assessment is introduced: fluency, or smooth and accurate reading (National Reading Panel, 2000). 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 free it up for the task of producing meaning (LaBerge & Samuels, 1974). Fluency in connected text pulls together the relationship described in the Simple View of Reading (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).

Using words read correctly per minute, researchers have shown that such scores predict later risk levels, 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 directly asking students comprehension questions after the reading, 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.

2.5.1. Oral Reading: Picture Book

All students taking the MAP Reading Fluency Adaptive Oral Reading tests interact with an onscreen "picture book" format, reading it aloud. 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. Students choose when to use the button to turn the 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 books, and all are narrative stories.

Table 2.23 presents the specifications for the Oral Reading: Picture Book measure. Table 2.24 presents descriptive data for each separate picture book, including word count and Lexile Framework® for Reading and Spache-Revised readability measures. All students who take the Adaptive Oral Reading forms get one of these books at their first task. Figure 2.19 presents a sample item.

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. The four indicators are 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 text, it is feasible to drop below 150L. In picture books, 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 books 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.

As a secondary measure of text complexity, the Spache-Revised index is reported. The Spache Readability Formula (Spache, 1953; Spache, 1974) was developed specifically for primary grade reading material (i.e., Grades K–3). The formula calculates the grade level of a text sample based on sentence length and the number of unfamiliar words.

Table 2.23. 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.

Picture Book	Lexile®	Spache-R	Word Count
Bear on the Bus	120L	1.2	59
Two Red Coats	130L	1.2	56
Water for Breakfast	150L	1.5	65
Jon Makes a Card	160L	1.2	61
Fred on a Hot Dog	160L	1.2	58
Walk Home with Best Friend	180L	1.2	47
Ken's Snow Day	190L	1.2	51
Duck in the Sink	210L	1.3	61
Why Why Why	220L	1.5	66
Jade's Grandma	230L	1.2	60
Star and Mom	310L	1.3	67

Table 2.24. Readability Measures and Word Count for Picture Book Texts

•	040	0	14	O = - 1	Decelline etc.	Distance	Deel
Flaure	Z.19.	Sample	item-	-Orai	Reading:	PICTURE	BOOK
		Jampio		••••			



2.5.2. 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 literal or low-inference comprehension questions. 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 WCPM and accuracy. Score for comprehension is percent correct out of 6. Figure 2.20 presents a sample passage, and Figure 2.21 presents a sample item containing passage comprehension questions associated with the passage.

Figure 2.20. Sample Passage







2.5.2.1. Passage Specifications

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 content, form, and tone. They were reviewed separately for any issues with bias or sensitivity.

In the first stage, two consecutive NWEA content specialists in primary grade literacy reviewed specific qualitative and quantitative criteria to select passages:

- The passage is well written and engaging.
- The passage is age appropriate for students in Grades K-3.
- 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 (i.e., levels of meaning or purpose; structure; language conventionality and clarity; and 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. Table 2.25 presents descriptive data for each passage used in MAP Reading Fluency.

Passage Title	Lexile®	Spache-R	Word Count
Art Bin	190L	1.1	148
Pink the Pig	200L	1.2	187
Mac the Cat	200L	1.2	157
Sal Gets Wet	210L	1.1	167
Ann's Bear	210L	1.2	187
Birds and Nests	220L	1.2	188
Zack in the Rain	220L	1.2	188
Be a Teacher	380L	2.0	207
Losing Teeth	400L	2.0	205
Bears	410L	2.0	200
Zoo	440L	2.0	204
A New Puppy	450L	1.9	201
Jay and Gus	460L	1.8	208
Airplanes	470L	2.0	206
Drinking Fountain	470L	3.8	212
Bing the Polar Bear	480L	1.8	207
Game Inventor	480L	3.0	205
Bus Love	490L	1.9	201
Old Photos	490L	2.8	221
Pam and the Toy Chest	490L	3.0	241
Butterflies and Moths	500L	2.9	216
Class Trip	500L	2.9	219
Spell Pizza	500L	2.9	236
Hamster on the Loose	500L	3.8	212
Playground Alien	500L	3.8	220
Blue Whales	520L	3.0	213
Popcorn Science	520L	4.0	215
Training a Puppy	540L	3.9	216
Bad Talent Show	560L	4.0	221
Emperor Penguins	560L	3.8	219
Ants	570L	2.8	213
Dad versus Socks	590L	3.8	221
Field Mice	610L	3.8	210

Table 2.25. Descriptive Metadata for Oral Reading Passages

2.5.2.2. Passage Comprehension Quiz

All questions were 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 questions. 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. Current item pool associated to each passage is six items, presented in a fixed order. The score is number correct out of 6, shown as a percentage.

Chapter 3: Item Development

For each component of early literacy included in MAP Reading Fluency, NWEA content specialists and external experts (i.e., professors and researchers with specialties in learning to read) reviewed relevant academic standards and progressions including the CCSS. Within these, 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. 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.

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 primary grades literacy for adherence to best practices for young students. The criteria used included the following:

- Each stem should clearly connect a student to the concept, idea, or skill being assessed.
- Each stem should clarify the functionality of the task, where necessary.
- Each stem should use simple, age-appropriate vocabulary.
- Each stem should use simple syntax, including features such as present tense, active voice, and short sentence length.
- Each stem should be worded positively and directly.

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

3.2. Item Writing and Review for Individual Items and Sets

Each item was written by NWEA content experts in primary grades literacy. Each received 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 choice 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 Grades K–3 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 Grades K–3 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.

Chapter 4: Test Administration

MAP Reading Fluency is administered through the NWEA Comprehensive Assessment Platform. Access is housed on the same platform as MAP® Growth[™] and MAP® Skills[™], 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. Students access the test application through the existing student dashboard login—the same URL and login currently used for MAP Skills student access. MAP Reading Fluency is best used seasonally, in conjunction with MAP Growth K–2. When students demonstrate the ability to read with near-perfect literal comprehension, they should transition to MAP Growth 2–5 tests and discontinue using MAP Reading Fluency.

MAP Reading Fluency is unique among measures of oral reading in that students read aloud into a microphone and are recorded and scored automatically by the speech scoring engine. Administration procedures can vary to accommodate a variety of student and educator needs.

4.1. 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 the Chromium[™] Browser (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 upto-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 will 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 requires 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.

4.2. 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 kindergarten or Grade 1, unless the student is testing on an iPad.

4.3. 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 inprogress test can be discarded, and the student will then 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.

4.4. Adaptive Presentation of Measures

Both the Adaptive Oral Reading and Foundational Skills formats present content adaptively. The Adaptive Oral Reading format follows a stage-adaptive methodology, as shown in Table 4.1. In Stage 1, all students read a picture book and complete the two-minute Sentence Reading Fluency measure. 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. Students performing below this threshold are presented instead with decoding and language comprehension measures. In addition, students who struggle with the comprehension questions after their first two passage attempts (=<66% correct) are presented with language comprehension measures. Table 4.2 summarizes the measures presented in the

Adaptive Oral Reading format and their reported outcomes. The Foundational Skills form presents the same measures, excluding the Oral Reading measures.

Stage 1	Stage 2	Stage 3
Oral Reading: Picture BookSentence Reading Fluency	Oral Reading: Passages and Comprehension Quiz	Language Comprehension
	Decoding Measures	Measures

 Table 4.1. Stages Presented Within the Adaptive Oral Reading Test Format

Table 4.2. Measures and Reported Outcomes

Domain	Measure	Code	Test Duration	Scoring Method	Reported Outcomes	
	Initial Sound Matching	001	2 minutes			
	Rhyming Words	015	2 minutes			
	Counting Syllables	017	1 minute			
Phonological	Onset-Rime Blending	018	1 minute	Dichotomously	Number correct and number	
Awareness	Phoneme Blending	019	1 minute	item level	attempted	
	Phoneme Counting	020	1 minute			
	Phoneme Addition/Deletion	021	2 minutes			
	Phoneme Substitution	022	2 minutes			
	Letter Knowledge	002	1 minute			
	Letter-Sound Fluency	003	1 minute			
	Decoding: CVC	007	1 minute		Number correct and number	
	Word Families: Initial Letter	023	1 minute	Dichotomously		
Phonics & Word Recognition	Build Words: One Letter	024	1 minute	scored at the		
recognition	Build Words: CVC	025	2 minutes	item level		
	Build Words: Single Syllable	026	2 minutes			
	Decoding: Single Syllable	027	1 minute			
	Sentence Reading Fluency	008	2 minutes			
Language	Listening Comprehension	004	Up to 30 seconds per item	Dichotomously	Number correct and number	
Comprehension	Picture Vocabulary	005	Up to 30 seconds per item	item level	attempted (typically all 15 are attempted)	
	Oral Reading: Picture Book	013	Up to 5 minutes	LanguaMetrics	WCPM; accuracy rate (words	
Oral Deading	Oral Reading: Passages	011	Up to 5 minutes	speech sconng software	time taken to read)	
Oral Reading	Oral Reading: Passage Comprehension Quiz	014	Up to 90 seconds per item	Dichotomously scored at the item level	Percent correct out of 6 for each quiz	

Chapter 5: Scoring & Reporting

5.1. Score Classification and Performance Levels

All student responses are scored automatically by the MAP Reading Fluency software. The reported outcomes of each measure in Table 4.2 constitute the raw scores for each measure. Oral Reading measures that yield WCPM 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. Raw scores are reported in the reporting site. In addition to the raw scores, one of the following performance levels is assigned to the results in each domain. Performance levels are color-coded as red, yellow, green, or blue:

- Exceeds Expectation: Blue
- Meets Expectation: Green
- Approaching Expectation: Yellow
- Below Expectation: Red

5.1.1. Raw Score Conversion to Performance Levels: Oral Reading Measures

Table 5.1 presents the minimum thresholds (i.e., minimum WCPM) for meeting expectation relative to grade-level text. Table 5.2 presents the performance level ranges based on WCPM for each grade and administration, which are drawn from published national norms (Hasbrouck & Tindal, 2017). Red, yellow, green, and blue color-coding is applied based on which quartile a student's score falls in. Students meet expectation if they read the minimum WCPM for a given grade and administration. 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.

	Minimum #Words Correct per Minute (WCPM)					
Grade	Fall Expectation Winter Expectation Spring Expectation					
K	No oral reading expected					
1	Up to 8 29 60					
2	50 84 100					
3	83 97 112					

Table 5.1. Exp	ectation Levels f	or Oral Reading	Fluency l	Based on WCP	M On Grade-level	Text
			, <i>,</i> .			

Table 5.2. Performance Levels based on Fall, Winter, and Spring WCPM

Grade	Performance Level	Fall WCPM	Winter WCPM	Spring WCPM			
	Exceeds Expectation		Any oral reading				
К	Meets Expectation	_	-	-			
	Approaching Expectation	_	-	-			
	Below Expectation	-	-	-			
1	Exceeds Expectation	9–140+	59–140+	91+			
	Meets Expectation	1–8	29–58	60–90			
	Approaching Expectation	-	16–28	34–59			
	Below Expectation	-	0–15	0–33			

Grade	Performance Level	Fall WCPM	Winter WCPM	Spring WCPM
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

5.1.2. Instructional Reading Levels

MAP Reading Fluency uses a combination of oral reading accuracy and passage understanding to report an instructional reading level range. Several research studies indicate that when early readers and less-proficient elementary grade students work with texts matched to their level, growth in reading fluency can be faster and engagement can be strengthened (e.g., O'Connor et al., 2002; Morris et al., 2011; see Allington, McCuiston, & Billen, 2015, for a more comprehensive review).

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, Fisher, and Frey (2016) note, readers "build muscle" in reading by working with more challenging texts. An instructional reading level is not inconsistent with this premise; instead, it indicates the *degree of instructional support* required to help students work with challenging 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).

Instructional reading level is computed based on independent, instructional, and frustrational criteria for comprehension and accuracy and is reported as a Lexile range. Instructional level comprehension is achieved with three or four correct responses out of six items presented. Independent criteria is five or six correct. For accuracy, instructional criteria is 88–94%, and independent criteria is 95% or more. The reported instructional reading level range is either a 100 Lexile range centered on a passage difficulty read at the instructional level, or a 100 Lexile range bounded on the lower end by an independently read text level, as shown in Table 5.3. For example, if a student reads a 400L passage with 94% accuracy and answers three out of six questions correctly about it, their instructional reading level will be 350–450L. If the 400L passage is read independently, the reported range is 400–500L. Students will read passages from different levels, and their performance across all scored passages is considered in reporting the range.

Table 5.3. Ex	pectations for	Instructional	Reading	Levels b	v Lexile
	poolationo ioi	mon averena			,

Grade	Lexile®
3	500L +
2	300L – 500L
1	Up to 300L

Table 5.4 presents the boundaries for performance levels for decoding accuracy, which is also classified, apart from its contribution to instructional reading level, according to ranges of percent accuracy on grade-level text.

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

Table 5.4.	Performance	Levels for I	Decoding	Accuracy	Based on	Percent	Accuracy
				· · · · · · · · · · · · · · · · · · ·			· · · · · · · · · · · · · · · · · · ·

Reading 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.

5.1.3. Raw Score Conversion to Performance Levels: Foundational Skills Measures

Foundational skills measures in MAP Reading Fluency are presented within the Foundational Skills test form or in the oral reading test form upon failure of sentence reading criteria to advance to oral reading. Foundational Skills include 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. These measures are presented adaptively based on student responses (i.e., number correct and percent correct). Each student moves through each of the two progressions based on their demonstrated ability. Performance levels are assigned at the level of the entire progression by comparing the observed zone of proximal development (ZPD) to grade-level expectations. ZPD levels are achievable from a series of related measures administered from each skill progression, as shown in Table 5.5 and Table 5.6. The ZPD level is highlighted in an onscreen representation of the progression, shown in the Individual Student Reports, and is stated in a narrative in the top summary section of the report.

Grade-level expectation is set at Level 1 in fall for kindergarten and Level 4 in winter for Grade 1. Table 5.5 and Table 5.6 present the pair of measures in each level within the Phonological Awareness and Phonics & Word Recognition domains, respectively. Students are administered 3–6 of these measures (typically 4–5) based on adaptive branching criteria in the test. From the administration and results of the several measures, a ZPD level is achieved, as outlined in Table 5.7. The Individual Student Report in Figure 5.2 shows this ZPD level outlined with a colored border, aligning to the performance level (Exceeds, Meets, Approaching, Below) for this performance for a given grade and term.

ZPD Level	Level 0: Rhymes & Syllables (Introduce)	Level 1: Rhymes & Syllables	Level 2: Initial Sounds	Level 3: Phoneme Blending & Segmenting	Level 4: Phoneme Manipulation	Level 5: Phoneme Manipulation (Reinforce)
Measures	Rhyming Words Measures phonological rhyme identification skills		Onset-Rime Blending Measures initial phoneme blending skills	Phoneme Blending Measures phoneme blending skills	Phoneme Addition/Deletion Measures phoneme manipulation skills	
Level	Counting Syllables Measures phonological syllable segmenting skills		Initial Sound Matching Measures initial phoneme identification skills	Phoneme Counting Measures phoneme segmenting skills	Phoneme Sub Measures phor skills	stitution neme manipulation

Table 5.6. ZPD Levels for Phonics & Word Recognition

ZPD Level	Level 0: Letters & Sounds (Introduce)	Level 1: Letters & Sounds	Level 2: Initial Letter & Word Families	Level 3: CVC Words	Level 4: One-Syllable words	Level 5: One- Syllable words (Reinforce)	
Measures	Letter-Sound Flue Measures letter sou correspondence kn	e ncy und oowledge	Build Words: One Letter Measures letter sound decoding skills in word	Decoding: CVC Measures early word decoding skills	ecoding: CVC leasures early word ecoding skills		
in Each Level	Letter Knowledge Measures letter identification 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 5.7. Performance Expectations by ZPD Level

			ZPD	Level			
Administration	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5	
Kindergarten							
Fall	Meets Exp	ectation	Exceeds Expectation				
Winter	Approaching Expectation		Meets Expectation	Exceeds Expectation			
Spring	Below Expectation		Approaching Expectation	Meets Expectation	Exceeds Exp	pectation	
Grade 1							
Fall	Below Exp	ectation	Approaching Expectation	Meets Expectation	pectation		
Winter	Below Exp	ectation		Approaching Expectation	ctation		
Spring	Below Exp	ectation			Approaching Expectation		
Grade 2							
Fall	Below Exp	ectation			Approaching	Expectation	
Winter	Below Exp	ectation					
Spring	Bolow Exp	ootation					
Grade 3							
Fall							
Winter	Below Exp	ectation					
Spring							

Oral language is assessed within the Foundational Skills section of the test and for any students in kindergarten and Grade 1 who proceed to passages but struggle to understand passages at the lowest Lexile levels. Listening Comprehension and Picture Vocabulary comprise the language comprehension section of the test. Each measure presents 15 items to the student, drawn randomly from a larger pool. Performance on each measure is classified and color-coded based on the number correct out of 15, as shown in Table 5.8.

		Number Corre	ect of 15	
Grade	Below Expectation	Approaching Expectation	Meets Expectation	Exceed Expectation
К	6 or less	78	9–11	12+
1	8 or less	9-11	12+	-
2	8 or less	9-11	12+	-
3	8 or less	9-11	12+	-

Table 5.8. Performance Expectations for Language Comprehension

5.2. Individual Student Reports

The Individual Student Report shows all raw scores achieved on a given assessment and includes a summary at the top with a skills profile and suggested 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. Figure 5.1 presents the report layout for a student who has read passages aloud and answered comprehension questions. Figure 5.2 presents the layout for students who have taken foundational skills measures within the Adaptive Oral Reading test format. This view is nearly identical when the Foundational Skills format is taken, with the exception of the picture book result that appears only when the oral reading form is taken. All other MAP Reading Fluency reports (e.g., Student Matrix view, Term Summary) are based on the data in these Individual Student Reports.



VORDS CORRECT PER MINUTE		INSTRUCTION	AL READING LEVEL		PROFILE & NEXT STEPS			
Meets grade level expectation. Islam meets grade level expectation. Fall expectation: 50 wcpm in 2nd grade:	text	A Appi Islan	oaching grade level expectation. reads at an instructional level between 1	70L and 270L.	Islam reads with good rate and un is insufficient. Building decoding accuracy, for rea	derstanding, but his/her decoding accuracy aders with good rate		
FEST DETAILS & RESULTS								
FEST DETAILS & RESULTS	LEXILE*	WCPM	DECODING ACCURACY	COMPREHENSION QUIZ	Z READABILITY			
FEST DETAILS & RESULTS Assage title e a Teacher	LEXILE* 380L	WСРМ 72	DECODING ACCURACY 88%	COMPREHENSION QUIZ	Z READABILITY Frustrational	Review Audio		
TEST DETAILS & RESULTS xxsAge time e a Teacher osing Teach	LEXILE* 380L 400L	WCPM 72 55	DECODING ACCURACY 88% 89%	COMPREHENSION QUIZ 67% 50%	Z READABILITY Frustrational Frustrational	© Review Audio © Review Audio		
IEST DETAILS & RESULTS ASSAGE TITLE le a Teacher osing Teeth rid Nests	LEXILE* 380L 400L 220L	WCPM 72 55 74	DECODING ACCURACY 88% 89% 93%	COMPREHENSION QUI 67% 50% 83%	Z READABILITY Frustrational Frustrational Instructional	Benter Audio Benter Audio Benter Audio Benter Audio Benter Audio		
REST DETAILS & RESULTS AND A CONTRACT OF A C	LEXILE* 380L 400L 220L	WCPM 72 55 74 WCPM	DECODING ACCURACY 88% 89% 93% DECODING ACCURACY	COMPREHENSION QUI 67% 50% 83%	Z READABILITY Frustrational Frustrational Instructional	Bentew Audio Bentew Audio Bentew Audio Bentew Audio		

Back to Student Matrix PERFORMANCE RESULTS, FOUNDATION Silvia Clester - 2nd grad	AL SKILLS (07/19/2018) e					
DECODING		LANGUAGE COM	MPREHENSION		PROFILE & NEXT STEPS	
Below grade level expectation. PHONOLOGICAL AWARENESS: Silvi Segmenting level (3) A Approaching grade level expectatic PHONICS/WORD RECOGNITION: Sil One-cyllable level (4)	a is working at the Blending & on. Wa is working at the Decodable:	M Meets LISTEN oral se M Meets PICTUR vocabu	grade level expectation. INC COMPREHENSION: Silvia understood 100% of com trences. grade level expectation. E VOCABULARY: Silvia matched pictures to 100% of or lary words.	nplex. al	Silvia has solid language comprehens words. He/she shows some awarenes Mapping obolics patterns to phonem Additional focus for students at risk	ion and can decode many one syllable ss of phonemes.
TEST DETAILS & RESULTS						
ICTURE BOOKS		WCPM	DECODING ACCURACY			
tar and Mom		NS	NS			Review Audio
ENTENCE READING FLUENCY @		RAW SCORE				
		12/15				
STENING COMPREHENSION @		RAW SCORE				
		15/15				
ICTURE VOCABULARY @		RAW SCORE				
ONE OF PROXIMAL DEVEL	OPMENT FOR FOUNDA					
HONOLOGICAL AWARNESS						
		reinforce with practice	ZPD	introduce wit	h support	
RHYMES & SYLLABLES	INITIAL SOUNDS		BLENDING & SEGMENTING	PHON	EME MANIPULATION	
RHYMING WORDS @	ONSET-RIME BLENDING		PHONEME BLENDING @	PHONEM	AE ADDITION/DELETION @	INSTRUCTIONAL RECOMENDATION Phoneme Segmentation Activities
COUNTING SYLLABLES @	INITIAL SOUND MATCHING	0	Raw Score: 10/12 PHONEME COUNTING @ Raw Score: 8/10	PHONE	Raw Score: 5/11 AE SUBSTITUTION @ Raw Score: 4/9	Segmenting & Blending Activities
HONICS/WORD RECOGNITION			reinforce with practi	ice	ZPD	
LETTER & SOUNDS	LETTERS IN WORDS		DECODABLE WORDS: CVC	DECO	OABLE WORDS: ONE SYLLABLE	
LETTER SOUND FLUENCY @	BUILD WORDS: ONE LETTE Raw Score:	R @ 15/15	DECODING: CVC @ Raw Score: 8/10	DECODI	NG: SINGLE SYLLABLE @ Raw Score: 7/7	INSTRUCTIONAL RECOMENDATION One Syllable Word Activities
LETTER KNOWLEDGE @	WORD FAMILIES: INITIAL L	TTER @	BUILDING WORDS: CVC @	BUILDIN	IG WORDS: SINGLE SYLLABLE 😡	

Communic Official and Domestic Forum deficience Obility Defi

5.2.1. Reader Profiles and Foundational Skills Profiles

A summary of student performance across all oral passage reading attempts is provided across three instructionally important dimensions of oral reading:

- Oral reading rate
- Decoding accuracy
- Passage comprehension •

Profile statements are generated for each complete test and are linked to suggested instructional next steps. Figure 5.3 presents an example Reader Profile that summarizes performance across the three sub-scores and 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, as shown in Figure 5.4.

Figure 5.3. Reader Profile

PROFILE & NEXT STEPS

Islam reads with good rate and understanding, but his/her decoding accuracy is insufficient.

Building decoding accuracy, for readers with good rate

Figure 5.4. Foundational Skills Profile

PROFILE & NEXT STEPS

Silvia has solid language comprehension and can decode many one syllable words. He/she shows some awareness of phonemes.

Mapping phonics patterns to phonemes

Additional focus for students at risk

Valencia et al. (2010) have shown that providing data on each of these components offers greater predictive validity than use of WCPM 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. The thresholds determining strength and difficulty are as follows, relative to grade-level text. Expectations for rate are adjusted up slightly for below grade-level text.

5.2.1.1. Oral Reading Rate

Oral reading rate is considered based on the expectation levels listed in Table 5.1. Increased instructional intensity is suggested for students reading at a rate significantly below expected levels. In addition, 70 WCPM is treated as a universal threshold below which weaker comprehension is a typical rather than unexpected outcome. Decoding is still largely effortful rather than automatic at these rates. This distinction affects which instructional foci and strategies are recommended for a student.

5.2.1.2. 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, McCuiston, and Billen (2015) find that a minimum of 95% accuracy predicted significant increases in both engagement and comprehension.

5.2.1.3. Passage Comprehension

Across all passages, five of six comprehension items answered correctly is used as a threshold for demonstrating basic understanding of the passage. Passage comprehension quiz scores above this threshold are treated as a necessary but insufficient indication of deeper, more subtle comprehension skills. This distinction is highlighted in the instructional guidance associated to some profiles. In all cases, the instructional foci and strategies recommended are evidencebased practices rooted in the inter-relationship between rate, accuracy, and comprehension.

5.3. 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 phonemelevel 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.

Chapter 6: Technical Characteristics

This section presents technical information on the measurement characteristics of the MAP Reading Fluency assessment based on Winter 2017, Fall 2017, Winter 2018, and Spring 2018 data. Since data collection for MAP Reading Fluency is an ongoing process, additional studies will be completed in the future and reported in subsequent technical reports.

MAP Reading Fluency includes a set of measures focusing on knowledge and skills with print, sounds, and the process of mapping print to sound (i.e., decoding). The major purposes of MAP Reading Fluency are to:

- 1. Point oral reading fluency data at immediate instructional decisions. These include finding appropriate instructional emphases and appropriate levels of text for individuals and for instructional groupings.
- 2. Offer one source of data for comparing a student's reading fluency to a general grade level expectation.

These purposes should be supported by reliability and validity evidence regarding score consistency, score purpose, and intended use. Motivated by these considerations, these studies sought to assess both reliability and validity information for the MAP Reading Fluency assessments. The major purposes of these studies are as follows:

- 1. Conduct classical item analyses for the decoding and language comprehension measures.
- 2. Provide reliability evidence for decoding and language comprehension measures.
- 3. Provide concurrent validity evidence for oral reading fluency measures by comparing oral reading scores to MAP Growth Reading scores.
- 4. Examine the relationship between passage oral reading fluency and sentence reading fluency.
- 5. Determine the effectiveness of sentence reading fluency in classifying students' oral reading fluency.
- 6. Evaluate the accuracy of the speech scoring engine for assessing oral reading fluency.

6.1. Data Collection

Data in this technical report are based on the Winter 2017 data collected from March 1–31, 2017; Fall 2017 data collected from September 1 – November 29, 2017; Winter 2018 data collected from December 1, 2017 – March 25, 2018; and Spring 2018 data collected from March 26 – July 20, 2018. All the item- and test-level analyses were conducted based on the Fall 2017, Winter 2018, and Spring 2018 data sets with the exception that the human and machine scoring reliability study and the classification accuracy study (see Section 6.7) were based on the Winter 2017 data. The targeted population was the U.S. national K–12 student population. However, the actual population used in the analyses was based on a convenience sample.

Table 6.1 – Table 6.4 present the sample sizes and percentages of students participating in the Winter 2017, Fall 2017, Winter 2018, and Spring 2018 tests, respectively, grouped by gender and ethnicity across grades. The results show that the sample was not exactly matched to the targeted population, but overall MAP Reading Fluency samples are close to the demographic characteristics of the U.S. national K–12 student population based on the U.S. Census Bureau's 2017 demographic data (U.S. Census Bureau, 2017).

				Gra	ade			
		К		1		2		3
Demographic	Ν	%	Ν	%	Ν	%	Ν	%
Gender								
Female	186	49.6	259	47.3	255	52.8	236	46.7
Male	189	50.4	288	52.7	228	47.2	269	53.3
Ethnicity								
American Indian or Alaskan	3	0.8	2	0.4	0	0.0	4	0.8
Asian	31	8.3	96	17.5	49	10.2	26	5.1
Black	45	12.0	41	7.5	48	9.9	49	9.7
Hispanic	50	13.4	52	9.5	35	7.2	83	16.4
Multiethnic	14	3.7	35	6.4	32	6.7	23	4.6
Not Specified or Other	15	4.0	11	2.0	15	3.1	10	2.0
White	217	57.8	310	56.7	304	62.9	310	61.4
Total	375	100.0	547	100.0	483	100.0	505	100.0

Table 6.1. Sample Population Demographics (Winter 2017)

Table 6.2. Sample Population Demographics (Fall 2017)

				Gra	ade			
		К	1	l	:	2	3	;
Demographic	N	N %		%	N	%	Ν	%
Gender								
Female	491	49.5	700	51.2	797	48.8	701	50.5
Male	501	50.5	666	48.8	837	51.2	688	49.5
Ethnicity								
American Indian or Alaskan Native	2	0.2	1	0.1	1	0.1	2	0.1
Asian	12	1.2	16	1.2	17	1.0	29	2.1
Black	219	22.1	230	16.8	250	15.3	224	16.1
Hispanic	210	21.2	301	22.0	289	17.7	209	15.0
Multiethnic	59	5.9	77	5.6	108	6.6	101	7.3
Native Hawaiian/Other Pacific Islander	_	-	2	0.1	2	0.1	6	0.4
Not Specified or Other	1	0.1	2	0.1	1	0.1	1	0.1
White	489	49.3	737	54.0	968	59.2	818	58.8
Total	992	100.0	1,366	100.0	1,636	100.0	1,390	100.0

Table 6.3. Sample Population Demographics (Winter 2018)

				Grade					
			К		1		2		;
	Demographic	Ν	%	N	%	N	%	Ν	%
Gender									
	Female	1,580	49.0	2,092	48.5	2,332	48.9	1,754	50.2
	Male	1,645	51.0	2,223	51.5	2,437	51.1	1,740	49.8

				ade					
	I	ĸ		1	2		:	3	
Demographic	N	%	N	%	N	%	N	%	
Ethnicity									
American Indian or Alaskan Native	21	0.7	22	0.5	29	0.6	28	0.8	
Asian	115	3.6	177	4.1	100	2.1	42	1.2	
Black	326	10.1	470	10.9	549	11.5	516	14.8	
Hispanic	287	8.9	411	9.5	569	11.9	318	9.1	
Multiethnic	124	3.8	116	2.7	151	3.2	93	2.7	
Native Hawaiian/Other Pacific Islander	8	0.2	13	0.3	4	0.1	6	0.2	
Not Specified or Other	444	13.8	616	14.3	248	5.2	193	5.5	
White	1,903	59.0	2,493	57.7	3,120	65.4	2,300	65.8	
Total	3,228	100.0	4,318	100.0	4,770	100.0	3,496	100.0	

Table 6.4. Sample Population Demographics (Spring 2018)

				Gra	ade			
	I	۲		1	:	2	:	3
Demographic	Ν	%	Ν	%	N	%	N	%
Gender								
Female	2,412	48.8	3,249	49.2	3,456	48.7	2,236	49.4
Male	2,530	51.2	3,354	50.8	3,634	51.3	2,290	50.6
Ethnicity								
American Indian or Alaskan Native	27	0.5	43	0.6	67	0.9	40	0.9
Asian	112	2.3	133	2.0	157	2.2	70	1.5
Black	465	9.4	679	10.2	645	9.1	379	8.3
Hispanic	878	17.7	1,301	19.6	1,465	20.6	697	15.3
Multiethnic	182	3.7	207	3.1	238	3.4	161	3.5
Native Hawaiian/Other Pacific Islander	12	0.2	13	0.2	7	0.1	7	0.2
Not Specified or Other	616	12.4	874	13.2	784	11.0	591	13.0
White	2,657	53.7	3,378	51.0	3,735	52.6	2,605	57.3
Total	4,949	100.0	6,628	100.0	7,098	100.0	4,550	100.0

6.2. Item Difficulty and Discrimination

All dichotomously scored MAP Reading Fluency items were analyzed for item difficulty and item discrimination based on classical test theory. Oral reading measures such as WCPM are scored in a score scale from 0 to around 200 and cannot be analyzed using classical item difficulty and discrimination analyses. Thus, Oral reading measures except Passage Comprehension Quiz (014), which is dichotomously scored, are excluded from these analyses.

In classical test theory, item difficulty is presented as the *p*-value (*p*) that shows the proportion of students who answer an item correctly. The *p*-value is bound by 0.0 and 1.0 and is derived by dividing the number of students who got the item correct by the total number of students who answered it. Item discrimination refers to the ability of an item to differentiate students who understand the concept being measured from those who do not (i.e., low-performing students vs. high-performing students). It is assessed by the correlation between how well students did on an item and how well they did on the entire test (i.e., their test score). The index used to measure item discrimination is a point-biserial correlation coefficient (r_{pbi}) for each item (*i*):

$$r_{pbi} = \frac{m_1 - m_x}{s_x} \sqrt{p_i / q_i}$$
(6.1)

where m_1 is the mean value on the continuous variable x (such as total test score) for students who answer the item correctly; m_x is the mean value on the continuous variable x for the entire group; s_x is the standard deviation; p_i is the proportion of students answering the item i correctly; and q_i is the proportion of students answering item i incorrectly.

Table 6.5, Table 6.6, and Table 6.7 present the mean *p*-values and point-biserial correlation coefficients across grades for Fall 2017, Winter 2018, and Spring 2018, respectively. The results indicate that the item difficulty of each measure generally decreases as the grade level increases. The mean item discriminations were higher than 0.2 with the exception of a low mean discrimination for Silent Sentence Reading: True/False (009) at Grade K and Grade 1 in Fall 2017. This could be due to the very young students not understanding the task near the beginning of the school year. The mean item discrimination of the same measure varies at different grade levels. A measure may be more discriminative at a particular grade level than at other grade levels. The results for Passage Comprehension Quiz (014) indicate that this measure has moderate mean item discrimination values and relatively easy items. The mean point-biserial correlations range from 0.31 to 0.50 and the mean *p*-values range from 0.53 to 0.80 across grades and administrations.

Overall, the items have desirable difficulty and reasonable discrimination. Besides measuring item quality, the item discrimination index can also be affected by other factors such as item difficulty (Lord, 1980, p. 33) and choice of group (i.e., the choice of high-ability vs. low-ability groups of students) separating the proportion of correct answers from incorrect.

			Testing	Grad	de K	Gra	de 1	Gra	de 2	Gra	de 3
Measure	Code	#Items	Time*	pm	r pbm	pm	r pbm	pm	r pbm	p m	r pbm
Initial Sound Matching	001	16	2	0.30	0.52	0.52	0.57	0.63	0.54	0.63	0.56
Letter Knowledge	002	20	1	0.71	0.63	0.93	0.37	0.96	0.28	0.96	0.28
Letter-Sound Fluency	003	20	1	0.60	0.61	0.87	0.38	0.92	0.30	0.91	0.28
Listening Comprehension	004	14	NS	0.72	0.49	0.88	0.49	0.92	0.50	0.93	0.50
Picture Vocabulary	005	14	NS	0.72	0.51	0.83	0.43	0.86	0.47	0.88	0.47
Decoding: CVC	007	20	1	0.32	0.39	0.66	0.52	0.75	0.39	0.77	0.43
Sentence Reading Fluency	008	30	2	0.33	0.36	0.65	0.50	0.89	0.38	0.95	0.23
Silent Sentence Reading: True/False**	009	32	2	0.54	0.08	0.63	0.10	0.81	0.24	0.89	0.23
Passage Comprehension Quiz	014	18	NS	0.69	0.41	0.79	0.42	0.77	0.38	0.77	0.31

Table 6.5. Mean *P*-Values (*p_m*) and Point-Biserial Correlation Coefficients (*r_{pbm}*) (Fall 2017)

*Testing time is in minutes. NS = not speeded.

**Silent Sentence Reading: True/False is no longer included on MAP Reading Fluency assessments.

Table 6.6. Mean <i>P</i> -Values (<i>pm</i>) and Point-Biseria	Correlation Coefficients (<i>r</i> _{pbm}) (Winter 2018)
--	--

			Testing	Grade K		Grade 1		Grade 2		Grade 3	
Measure	Code	#Items	Time*	p _m	r pbm	p _m	r pbm	p _m	r pbm	pm	r pbm
Initial Sound Matching	001	16	2	0.48	0.57	0.61	0.56	0.61	0.59	0.61	0.60
Letter Knowledge	002	20	1	0.90	0.44	0.96	0.26	0.96	0.30	0.95	0.43
Letter-Sound Fluency	003	20	1	0.82	0.46	0.91	0.28	0.91	0.31	0.90	0.35
Listening Comprehension	004	21	NS	0.78	0.46	0.88	0.48	0.90	0.47	0.90	0.59

			Testing	Gra	de K	Gra	de 1	Gra	de 2	Gra	de 3
Measure	Code	#Items	Time*	p m	r pbm	pm	r pbm	pm	r pbm	p m	r pbm
Picture Vocabulary	005	20	NS	0.84	0.51	0.88	0.51	0.89	0.53	0.91	0.54
Decoding: CVC	007	20	1	0.47	0.48	0.80	0.45	0.90	0.38	0.93	0.35
Sentence Reading Fluency	008	45	2	0.38	0.27	0.74	0.39	0.88	0.30	0.91	0.24
Passage Comprehension Quiz	014	18	NS	0.53	0.43	0.70	0.43	0.73	0.40	0.67	0.50
Rhyming Words	015	30	2	0.56	0.53	0.68	0.49	0.66	0.52	0.65	0.51
Counting Syllables	017	20	2	0.42	0.49	0.51	0.54	0.55	0.52	0.59	0.54

*Testing time is in minutes. NS = not speeded.

Table 6.7. Mean *P*-Values (*p_m*) and Point-Biserial Correlation Coefficients (*r_{pbm}*) (Spring 2018)

			Testing	Grade K		Grade 1		Grade 2		Grade 3	
Measure	Code	#Items	Time*	pm	r pbm	p m	r pbm	p m	r pbm	pm	r pbm
Initial Sound Matching	001	16	2	0.52	0.59	0.61	0.56	0.61	0.57	0.63	0.56
Letter Knowledge	002	20	1	0.92	0.37	0.96	0.26	0.96	0.25	0.96	0.35
Letter-Sound Fluency	003	20	1	0.86	0.40	0.91	0.29	0.91	0.27	0.90	0.32
Listening Comprehension	004	36	NS	0.74	0.47	0.82	0.46	0.86	0.40	0.85	0.51
Picture Vocabulary	005	35	NS	0.82	0.52	0.87	0.53	0.89	0.48	0.88	0.60
Decoding: CVC	007	36	1	0.62	0.38	0.80	0.35	0.75	0.30	0.71	0.38
Sentence Reading Fluency	008	45	2	0.45	0.30	0.81	0.37	0.92	0.24	0.94	0.20
Passage Comprehension Quiz	014	18	NS	0.70	0.30	0.77	0.37	0.80	0.40	0.79	0.49
Rhyming Words	015	30	2	0.58	0.55	0.65	0.53	0.66	0.53	0.64	0.62
Counting Syllables	017	20	1	0.39	0.50	0.45	0.53	0.49	0.52	0.51	0.57
Onset-Rime Blending	018	29	1	0.82	0.38	0.89	0.31	0.90	0.23	0.90	0.29
Phoneme Blending	019	30	1	0.60	0.47	0.73	0.38	0.75	0.36	0.75	0.43
Phoneme Counting	020	30	1	0.46	0.51	0.58	0.56	0.60	0.56	0.59	0.51
Phoneme Addition/Deletion	021	30	2	0.49	0.38	0.59	0.40	0.65	0.41	0.61	0.42
Phoneme Substitution	022	30	2	0.43	0.38	0.51	0.39	0.53	0.36	0.56	0.41
Word Families: Initial Letter	023	20	1	0.72	0.51	0.85	0.40	0.86	0.37	0.84	0.37
Build Words: One Letter	024	30	1	0.79	0.49	0.90	0.36	0.90	0.28	0.90	0.33
Build Words: CVC	025	30	2	0.70	0.61	0.80	0.46	0.80	0.46	0.77	0.48
Build Words: Single Syllable	026	30	2	0.32	0.58	0.52	0.50	0.56	0.48	0.53	0.54
Decoding: Single Syllable	027	30	1	0.41	0.34	0.62	0.39	0.69	0.42	0.67	0.47

*Testing time is in minutes. NS = not speeded.

6.3. Summary Raw Score Statistics

Table 6.8, Table 6.9, and Table 6.10 present summary raw score statistics for Fall 2017, Winter 2018, and Spring 2018, respectively, including the number of responses, the mean score, the standard deviation (SD) of the scores, and the allowed testing time in minutes for speeded measures. Once again, the oral reading measures except Passage Comprehension Quiz (014) are not included in these analyses.

For each grade, the number of responses does not match the sample sizes presented in Section 6.1 because the measures were included in different test forms. Most measures were included in the Foundational Skills forms only, but Passage Comprehension Quiz (014) was included in the Adaptive Oral Reading form and Sentence Reading Fluency (008) was included in both the Foundational Skills and Adaptive Oral Reading forms. Thus, the number of responses for each measure depended on the number of students who took each form. In general, the mean score of each measure increases as the grade level increases.

Table 6.8. Summary Raw Score Statistics (Fall 2017)

			Testing	ng Grade K			Grade 1			Grade 2			Grade 3		
Measure	Code	#Items	Time [*]	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Initial Sound Matching	001	16	2	939	3.31	2.77	902	5.64	3.44	617	6.35	3.35	207	6.41	3.65
Letter Knowledge	002	20	1	941	10.58	5.43	902	15.23	3.94	617	16.6	3.60	206	16.96	3.64
Letter-Sound Fluency	003	20	1	937	6.59	4.32	903	11.17	3.47	617	12.23	3.22	206	12.28	3.24
Listening Comprehension	004	14	NS	946	9.81	3.43	938	12.2	2.32	715	12.84	1.95	451	12.96	1.81
Picture Vocabulary	005	14	NS	945	9.88	3.03	937	11.48	2.07	715	11.99	1.76	451	12.21	1.77
Decoding: CVC	007	20	1	963	4.45	3.40	1,098	9.13	5.85	617	10.82	5.11	206	12.26	5.24
Sentence Reading Fluency	008	30	2	477	7.75	4.75	899	11.63	6.62	1,439	16.5	6.98	1,348	21.37	6.27
Silent Sentence Reading: True/False	009	32	2	124	13.65	4.84	767	11.35	5.77	1,432	12.6	6.18	1,343	16.22	6.15
Passage Comprehension Quiz	014	18	NS	25	11.08	5.12	218	13.33	4.58	850	13.65	3.82	1,154	13.28	4.46

*Testing time is in minutes. NS = not speeded.

Table 6.9. Summary Raw Score Statistics (Winter 2018)

			Testing		Grade K			Grade 1			Grade 2			Grade 3	
Measure	Code	#Items	Time [*]	N	Mean	SD	N	Mean	SD	N	Mean	SD	Ν	Mean	SD
Initial Sound Matching	001	16	2	2,738	4.83	3.32	2,410	5.99	3.36	1,024	6.18	3.55	362	6.63	3.76
Letter Knowledge	002	20	1	2,741	14.34	4.46	2,412	16.73	3.52	1,021	17.13	3.51	361	17.46	3.56
Letter-Sound Fluency	003	20	1	2,737	10.18	4.05	2,414	12.41	3.39	1,027	12.43	3.38	360	12.22	3.87
Listening Comprehension	004	21	NS	1,507	10.86	2.65	2,420	12.25	2.14	1,028	12.55	1.95	62	12.61	2.45
Picture Vocabulary	005	20	NS	1,509	11.74	2.31	2,418	12.37	1.99	1,028	12.40	2.01	62	12.71	1.99
Decoding: CVC	007	20	1	2,986	5.94	4.46	4,057	13.20	5.57	4,420	16.99	3.98	3,298	18.24	2.72
Sentence Reading Fluency	008	45	2	2,257	8.40	5.62	4,010	14.35	7.35	4,403	20.45	7.23	3,297	24.28	7.16
Passage Comprehension Quiz	014	18	NS	301	9.06	4.93	1,717	12.51	3.16	3,472	13.00	3.35	2,998	14.27	3.75
Rhyming Words	015	30	2	2,735	6.26	3.84	2,410	7.82	3.88	1,026	7.84	4.10	360	8.39	4.26
Counting Syllables	017	20	2	2,738	7.86	4.76	2,420	9.57	5.45	1,026	10.32	5.26	361	11.17	5.48

*Testing time is in minutes. NS = not speeded.

			Testing	Grade K		Grade 1			Grade 2			Grade 3			
Measure	Code	#Items	Time*	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Initial Sound Matching	001	16	2	4,299	4.82	3.25	3,397	5.76	3.29	1,279	5.95	3.28	371	6.39	3.47
Letter Knowledge	002	20	1	4,311	14.65	4.38	3,400	16.50	3.70	1,278	16.79	3.49	370	16.98	3.78
Letter-Sound Fluency	003	20	1	4,312	10.71	3.78	3,406	12.38	3.28	1,270	12.32	3.09	370	12.06	3.41
Listening Comprehension	004	36	NS	3,907	10.33	2.79	3,322	11.49	2.33	1,260	12.00	1.94	365	11.85	2.45
Picture Vocabulary	005	35	NS	3,909	11.53	2.63	3,320	12.14	2.38	1,257	12.39	2.10	365	12.35	2.54
Decoding: CVC	007	36	1	4,306	7.74	4.33	3,399	11.63	4.72	1,278	18.19	8.72	369	15.91	7.49
Sentence Reading Fluency	008	45	2	4,052	8.50	5.68	6,500	14.89	6.88	7,082	20.58	6.79	4,545	24.33	7.17
Passage Comprehension Quiz	014	18	NS	629	11.16	4.51	3,213	14.24	3.07	5,816	14.23	3.05	4,178	15.41	3.21
Rhyming Words	015	30	2	1,815	5.92	3.64	1,611	6.73	3.78	606	7.41	3.87	181	7.73	4.61
Counting Syllables	017	20	1	2,165	7.07	4.58	1,714	8.27	5.04	621	9.03	5.08	188	9.59	5.51
Onset-Rime Blending	018	29	1	2,055	9.11	3.97	1,713	10.71	3.86	649	10.96	3.97	177	10.99	4.27
Phoneme Blending	019	30	1	2,056	4.49	2.74	1,710	5.96	2.80	649	5.90	2.84	181	6.07	3.09
Phoneme Counting	020	30	1	2,205	5.02	3.20	1,710	6.24	3.61	631	6.57	3.76	186	6.50	3.68
Phoneme Addition/Deletion	021	30	2	2,139	4.19	2.35	1,690	5.05	2.62	659	5.79	2.83	183	5.72	2.63
Phoneme Substitution	022	30	2	2,176	3.49	2.06	1,691	4.03	2.24	616	4.43	2.25	192	4.80	2.49
Word Families: Initial Letter	023	20	1	2,225	5.82	3.00	1,693	8.43	3.02	629	9.16	3.16	190	9.16	3.25
Build Words: One Letter	024	30	1	2,107	1.64	1.39	1,665	1.90	1.43	651	2.01	1.14	186	2.18	1.15
Build Words: CVC	025	30	2	2,167	6.92	4.65	1,715	10.20	4.90	620	11.43	4.99	189	11.94	5.40
Build Words: Single Syllable	026	30	2	2,149	3.03	2.97	1,697	6.03	3.73	658	7.32	3.93	186	7.59	4.87
Decoding: Single Syllable	027	30	1	2,940	5.00	3.77	3,398	8.68	4.74	92	11.75	6.57	112	11.36	5.99

Table 6.10. Summary Raw Score Statistics (Spring 2018)

*Testing time is in minutes. NS = not speeded.

6.4. Reliability Evidence

6.4.1. Overview

Reliability is the degree to which scores remain consistent over an assessment procedure (Nitko, 2004), or the extent to which an assessment would likely yield the same results if it were re-administered. The consistency of scores can be represented in numerous ways. For example, internal consistency assesses how items function within a test; test-retest reliability measures how items function over time; an alternative forms method investigates how items function in two parallel forms; and rater reliability looks at how independent human scorers score common items. This section focuses on internal consistency reliability.

The internal consistency of a test investigates the stability of scores from one sample of content to another. One approach is to split all test items into two groups and then correlate student scores on the two half-tests. Another commonly used approach is Cronbach's alpha, which is a function of the number of test items and the average inter-correlation among the items. However, these traditional methods depend on all test takers taking a common test consisting of the same test items. For some MAP Reading Fluency measures, items were randomly administered to students and not all students taking a common test were administered the same items. Thus, application of these methods to MAP Reading measures is not appropriate.

An equally valid alternative, the marginal reliability coefficient (Samejima, 1977, 1994), is widely used to estimate the internal consistency reliability for adaptive tests. This coefficient incorporates measurement error as a function of the test score. In effect, it is the result of combining measurement error estimated at different points on the achievement scale into a single index. This method of calculating internal consistency, ρ_{θ} , yields results that are nearly identical to coefficient alpha when both methods are applied to the same fixed-form tests. The approach taken is given by:

$$\rho_{\theta} = \frac{\sigma_{\theta}^2 - M_{S_{\theta}^2}}{\sigma_{\theta}^2} \tag{6.2}$$

where σ_{θ}^2 is the observed variance of the achievement estimates, θ , and $M_{S_{\theta}^2}$ 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. To calculate the marginal reliability coefficients, each measure was calibrated using the Rasch model (Rasch, 1960/1980).

6.4.2. Summary Marginal Reliability Statistics

Table 6.11 presents the marginal reliability coefficients of the Phonological Awareness, Phonics & Word Recognition, and Language Comprehension domain scores of the Foundational Skills forms by administration and grade. The marginal reliabilities for Phonics & Word Recognition are higher than 0.8 across all grades and administrations. The marginal reliabilities of Phonological Awareness are higher than 0.8 except for Fall 2017 because only one measure, Initial Sound Matching (001), was developed for this domain at that time. The marginal reliabilities of Language Comprehension are lower than the other two domains across all grades and administrations because this domain only had two measures, which is much fewer than the other two domains.

The measures included in each domain are short, speeded tests with the exceptions of measures 004, 005, and 014 that are not speeded. The testing time for each measure is often only 1 or 2 minutes. Factors such as test speed and limited number of items administered within such a short amount of time can lower the reliability of an individual measure and thus lower the reliability of domain scores. Performance level scores at the domain level will likely demonstrate higher reliability. Reliability of performance level scores will be evaluated using 2018–2019 student response data and reported in a subsequent version of this technical report.

Administration	Domain*	Grade K	Grade 1	Grade 2	Grade 3
	Phonological Awareness	0.53	0.69	0.66	0.62
Fall 2017	Phonics & Word Recognition	0.89	0.88	0.87	0.85
	Language Comprehension	0.81	0.62	0.46	0.53
	Phonological Awareness	0.86	0.88	0.88	0.89
Winter 2018	Phonics & Word Recognition	0.86	0.84	0.84	0.85
	Language Comprehension	0.70	0.57	0.55	0.64
	Phonological Awareness	0.86	0.87	0.85	0.88
Spring 2018	Phonics & Word Recognition	0.92	0.90	0.89	0.91
	Language Comprehension	0.75	0.66	0.57	0.67

Table 6.11. Summary Marginal Reliabilities for Foundational Skill Forms (Fall 2017, Winter 20	018,
and Spring 2018)	

*The measures included in each domain are slightly different across administrations because of changes in the test design and development. Below is a list of the measures included in each domain of each administration.

Fall 2017:

- Phonological Awareness: 001
- Phonics & Word Recognition: 002,003,007,008,009
- Language Comprehension: 004, 005

Winter 2018:

- Phonological Awareness: 001, 015,017
- Phonics & Word Recognition: 002,003,007,008
- Language Comprehension: 004, 005

Spring 2018:

- Phonological Awareness: 001, 015,017,018,019,020,021,022
- Phonics & Word Recognition: 002,003,007,023,024,025,026,027,008
- Language Comprehension: 004, 005

6.5. Concurrent Validity Evidence

Concurrent validity refers to the extent to which an assessment's test scores predict student performance on other assessments of the same content area. It answers the question, "How well do the scores from this test correspond to the scores from an established test that references some other scale in the same content area?" It is calculated based on the total test scores of both assessments and expressed as a Pearson product-moment correlation coefficient with a value between 0.0 and 1.0.

Concurrent validity evidence for MAP Reading Fluency was determined between the WCPM scores from Oral Reading: Passages (011), the scores from Passage Comprehension Quiz (014), and MAP Growth Reading Rasch Unit (RIT) scores (including MAP Growth K–2). The passage WCPM scores measure reading fluency, whereas the passage comprehension scores and MAP Growth scores measure reading achievement. Students' MAP Reading Fluency

scores were matched to their MAP Growth scores obtained from the same administration (e.g., MAP Reading Fluency scores from Fall 2017 were matched to MAP Growth scores from Fall 2017). Only data from students who took both MAP Reading Fluency and MAP Growth in the same administration were used in this analysis.

Table 6.12, Table 6.13, and Table 6.14 present the results for Fall 2017, Winter 2018, and Spring 2018, respectively. The results show moderate Pearson product-moment correlation coefficients (*r*) between MAP Reading Fluency passage WCPM and MAP Growth Reading scores across administrations and grades. Similarly, the relationships between passage comprehension and MAP Growth Reading scores are generally moderate. However, some passages from Winter 2018 and Spring 2018 have correlations lower than 0.3. This may be due to the small sample sizes of these passages. The moderate relationships reveal that students' MAP Reading Fluency scores and their MAP Growth scores are related to some extent but are not highly related as these two tests are designed to measure slightly different constructs.

			WCF MAP Gro	PM & owth RIT	Compreh MAP Gro	ension & owth RIT
Grade	Passage Title	Passage Code	N	r	N	r
	Losing Teeth	P1112	12	0.55	16	0.50
ĸ	Bears	P1111	10	0.50	15	0.57
K	Sal Gets Wet	P0111	11	0.58	18	0.37
	Pink the Pig	P0112	9	0.60	16	0.46
	Losing Teeth	P1112	79	0.66	105	0.55
1	Bears	P1111	83	0.65	110	0.49
I	Sal Gets Wet	P0111	63	0.53	92	0.43
	Pink the Pig	P0112	53	0.34	87	0.40
	Bears	P1111	561	0.59	691	0.44
	Be a Teacher	P1116	122	0.52	148	0.53
n	Old Photos	P2111	379	0.60	487	0.49
Z	Blue Whales	P2116	125	0.63	156	0.52
	Emperor Penguins	P3117	213	0.64	300	0.44
	Pink the Pig	P0112	121	0.49	205	0.37
	Zoo	P1114	32	0.56	46	0.60
	A New Puppy	P1113	170	0.47	248	0.32
	Airplanes	P1115	333	0.58	384	0.35
	Old Photos	P2111	639	0.59	753	0.49
3	Butterflies and Moths	P2113	234	0.63	277	0.48
	Class Trip	P2117	182	0.57	218	0.41
	Field Mice	P3112	161	0.61	186	0.37
	Bad Talent Show	P3111	457	0.60	550	0.54
	Training a Puppy	P3116	168	0.65	208	0.55

 Table 6.12. Pearson Product-Moment Correlation Coefficients (r) between MAP Reading Fluency

 Passage WCPM and Comprehension Scores vs. MAP Growth Scores (Fall 2017)

			WCPM & MAP Growth RIT		Comprehe MAP Gro	ension & wth RIT
Grade	Passage Title	Passage Code	N	r	N	r
	Zoo	P1114	33	0.41	43	0.32
	A New Puppy	P1113	18	0.38	22	0.58
IZ.	Ann's Bear	P0113	37	0.24	104	0.69
ĸ	Bird Nests	P0114	28	0.56	92	0.67
	Sal Gets Wet	P0111	42	0.58	165	0.70
	Pink the Pig	P0112	53	0.14	162	0.61
	Zoo	P1114	514	0.45	607	0.48
	A New Puppy	P1113	510	0.34	624	0.36
	Bus Love	P1118	274	0.42	399	0.49
	Jay and Gus	P1119	155	0.24	219	0.44
4	Ants	P2114	212	0.34	250	0.42
1	Baker Brother	P2115	144	0.14	163	0.47
	Ann's Bear	P0113	295	0.36	407	0.45
	Bird Nests	P0114	138	0.26	207	0.47
	Sal Gets Wet	P0111	225	0.30	369	0.42
	Pink the Pig	P0112	339	0.39	450	0.38
	Zoo	P1114	555	0.51	695	0.35
	A New Puppy	P1113	779	0.55	990	0.38
	Bus Love	P1118	472	0.61	602	0.46
	Jay and Gus	P1119	251	0.67	312	0.43
	Ants	P2114	1,116	0.57	1,278	0.41
2	Baker Brother	P2115	1,575	0.62	1,847	0.50
2	Spell Pizza	P2119	488	0.62	623	0.56
	Pam and the Toy Chest	P2118	470	0.57	620	0.49
	Drinking Fountain	P3114	495	0.66	563	0.38
	Dad Versus Socks	P3115	760	0.56	878	0.52
	Ann's Bear	P0113	504	0.56	644	0.41
	Bird Nests	P0114	222	0.58	322	0.41
	Ants	P2114	206	0.37	357	0.36
	Baker Brother	P2115	210	0.54	377	0.46
	Spell Pizza	P2119	349	0.60	463	0.63
З	Pam and the Toy Chest	P2118	353	0.57	446	0.48
5	Drinking Fountain	P3114	2,085	0.62	2,410	0.48
	Dad Versus Socks	P3115	1,954	0.60	2,276	0.56
	Playground Alien	P3118	712	0.60	893	0.57
	Popcorn Science	P3119	703	0.57	910	0.57

Table 6.13. Pearson Product-Moment Correlation Coefficients (r) between MAP Reading FluencyPassage WCPM and Comprehension Scores vs. MAP Growth Scores (Winter 2018)

			WCP MAP Gro	M & wth RIT	Comprehe MAP Gro	ension & wth RIT
Grade	Passage Title	Passage Code	N	r	N	r
	Bus Love	P1118	29	0.43	34	0.55
	Jay and Gus	P1119	20	0.09	24	-0.03
	Losing Teeth	P1112	49	0.24	58	0.36
K	Bears	P1111	36	0.70	43	0.33
	Ann's Bear	P0113	196	0.37	393	0.64
	Bird Nests	P0114	162	0.28	351	0.58
	Zack in the Rain	P0115	221	0.32	452	0.66
	Bus Love	P1118	522	0.47	640	0.44
	Jay and Gus	P1119	509	0.50	620	0.45
	Losing Teeth	P1112	567	0.47	680	0.50
	Bears	P1111	530	0.50	630	0.45
	Old Photos	P2111	381	0.49	452	0.47
1	Butterflies and Moths	P2113	446	0.44	532	0.44
	Class Trip	P2117	202	0.50	226	0.26
	Game Inventor	P2112	343	0.45	417	0.50
	Ann's Bear	P0113	199	0.34	326	0.41
	Bird Nests	P0114	415	0.40	617	0.47
	Zack in the Rain	P0115	1,873	0.47	2,570	0.38
	Bus Love	P1118	1,354	0.49	1,702	0.50
	Jay and Gus	P1119	1,249	0.57	1,595	0.47
	Losing Teeth	P1112	444	0.27	557	0.39
	Bears	P1111	168	0.41	261	0.43
	Old Photos	P2111	1,100	0.51	1,258	0.48
	Butterflies and Moths	P2113	1,081	0.56	1,232	0.49
2	Class Trip	P2117	1,157	0.54	1,300	0.35
2	Game Inventor	P2112	1,142	0.53	1,279	0.54
	Field Mice	P3112	900	0.52	1,039	0.40
	Hamster on the Loose	P3113	646	0.51	721	0.29
	Bad Talent Show	P3111	784	0.47	895	0.41
	Ann's Bear	P0113	1,050	0.52	1,279	0.39
	Bird Nests	P0114	1,017	0.51	1,300	0.42
	Emperor Penguins	P3117	345	0.55	788	0.41
	Losing Teeth	P1112	802	0.57	930	0.45
	Bears	P1111	803	0.59	949	0.46
	Old Photos	P2111	113	0.43	181	0.32
	Butterflies and Moths	P2113	92	0.55	149	0.46
	Class Trip	P2117	219	0.54	302	0.37
3	Game Inventor	P2112	181	0.50	257	0.38
0	Field Mice	P3112	800	0.62	902	0.49
	Hamster on the Loose	P3113	838	0.62	945	0.48
	Bad Talent Show	P3111	830	0.63	949	0.56
	Playground Alien	P3118	2,131	0.61	2,360	0.46
	Popcorn Science	P3119	2,084	0.58	2,326	0.46
	Emperor Penguins	P3117	345	0.58	930	0.49

 Table 6.14. Pearson Product-Moment Correlation Coefficients (r) between MAP Reading Fluency

 Passage WCPM and Comprehension Scores vs. MAP Growth Scores (Spring 2018)

6.6. Relationship between Sentence Reading Fluency and Oral Reading Fluency

Because stronger readers' comprehension is highly correlated to sentence reading fluency, students with strong sentence reading fluency skills are likely to read with good oral reading fluency. An important feature of MAP Reading Fluency is that it combines data on a student's decoding accuracy and comprehension with their oral reading rate to generate a profile of strengths and needs in oral passage reading. The relationship between the Sentence Reading Fluency (008) and Oral Reading: Passages (011) measures is

the foundation on which to combine both decoding and comprehension to establish a more comprehensive picture of a student's reading ability.

Table 6.15 presents the Pearson product-moment correlation coefficients (*r*) between WCPM (human scores) of passage oral reading fluency and sentence reading fluency across passages and grades. Human scores were used to allow the full range of scores that will serve as the gold standard for the WCPM true score. The results show that both measures have moderate to high correlations across different passages and grades.

				Passage	WCPM	SRF Raw	Score*	
Grade	Passage Title	Passage Code	Ν	Mean	SD	Mean	SD	r
K	Sal Gets Wet	0111	33	50.2	34.8	12.0	6.6	0.91
n	Pink the Pig	0112	26	51.7	31.0	14.4	6.7	0.84
1	Bears	1111	181	86.5	34.3	21.5	6.3	0.85
I	Losing Teeth	1112	176	94.5	40.7	22.6	5.1	0.74
	Old Photos	2111	336	103.0	36.1	25.0	4.7	0.72
2	Game Inventor	2112	305	102.3	36.4	25.1	4.6	0.73
	Butterflies and Moths	2113	316	98.5	34.0	24.8	5.1	0.74
	Bad Talent Show	3111	321	109.9	36.1	26.2	3.9	0.66
3	Field Mice	3112	332	107.7	32.0	26.5	3.6	0.62
	Hamster on the Loose	3113	320	111.1	36.3	26.1	4.3	0.64

 Table 6.15. Pearson Product-Moment Correlation Coefficients (r) between Sentence Reading

 Fluency and Passage Oral Reading Fluency (Winter 2017)

*SRF = Sentence Reading Fluency

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

While the Sentence Reading Fluency measure is a strong predictor of WCPM for students and can be used to classify students in oral reading fluency, it is important to know the accuracy of this classification. The effectiveness of using sentence reading fluency to classify oral reading fluency was investigated using logistic regression. Sentence reading fluency was the independent variable, and oral reading fluency was the dependent variable. In general, logistic regression can be used to investigate the relationship between discrete responses (e.g., binary, ordinal, and nominal responses) and a set (vector) of discrete or continuous independent variables x.

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

$$logit(p) = log\left(\frac{p}{1-p}\right) = \alpha + \beta' x$$
(6.3)

where *p* is the response probability modeled, α is the intercept parameter, and $\beta = (\beta_1, ..., \beta_k)'$ is the vector of *k* slope parameters.

The purpose of a logistic regression is to find the most parsimonious set of predictors that are most effective in predicting the dependent variable. The null and alternative hypotheses for assessing overall logistic regression model fit are as follows:

- H_0 : The model fits the data.
- H_a : The model does not fit the data.

Non-rejection of the H_0 means the independent variables are statistically significant from zero or they improve model fit. However, this does not guarantee that the predictors or independent variables are practically significant.

The effectiveness of sentence reading fluency in classifying oral reading fluency was investigated with two approaches.

- 1. Approach 1: Sentence reading fluency was treated as a continual variable and oral reading fluency as a dichotomous variable.
- 2. Approach 2: Both sentence reading fluency and oral reading fluency were treated as dichotomous variables by using cutoff values.

The cutoff value is what converts a continuous measure to a dichotomous measure that, in turn, is used to classify observations. The cutoff values were evaluated using a combination of receiver operating characteristic (ROC) curve analysis and content experts' judgements. After the ROC curve analysis, the cutoff values for both sentence reading fluency and oral reading fluency were determined. The cutoff value for the oral reading fluency was 30 WCPM, and the cutoff values for sentence reading fluency were multiple and depended on the grade. The terms used for the classification accuracy included sensitivity, specificity, false positive rate, false negative rate, base rate, and overall classification accuracy, as shown in Figure 6.1.

	Observed Performance of Oral Reading Fluency								
		Good	Poor	Total					
Predicted Performance by Sentence Reading Fluency	Good	True Positive (TP)	False Positive (FP)	TP+FP					
	Poor	False Negative (FN)	True Negative (TN)	FN+TN					
	Total	TP+FN	FP+TN	TP+FP+FN+TN					
FPR=False Positive Rate=[FP/(F	P + 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)]									

Elaura 6 1	Classifiantian	of Oral	Dooding	Eluonov	hu	Contonoo	Dooding	Eluanov
FIGULE 0.1.	Classification	UI UI aI	Reaulinu	FILLETICV	. DV 4	Sentence	Reautitu	FILLETICV

In Approach 1, the predictor was a continual variable. Table 6.16 presents the odds ratio results by logistic regression across passages and grades. The results show that sentence reading fluency is significant in the fitted model for all passages except for Passage 3112, in which estimation was not converged. Overall, the range of odds ratios for all passages is from low 1.36 to high 2.95. Because the odds ratio is a measure of the odds that good performance in oral reading fluency will occur given good performance predicted by sentence reading fluency, compared to the odds of good performance in oral reading fluency given poor performance predicted by sentence reading fluency, the results in Table 6.16 represent the log odds of the oral reading fluency per unit increase in the value of sentence reading fluency. For example, for Passage 2111, the odds of good as defined by the cut point value performance in oral reading fluency is estimated between 1.45 and 3.64 times for 1 score increase in sentence reading fluency. All odds ratio values presented in Table 6.16 are larger than 1, which indicates that sentence reading fluency is associated with higher odds of oral reading fluency.

Grade	Passage Title	Passage Code	Lower CI*	Upper CI*	Odds Ratio
K	Sal Gets Wet	0111	1.20	2.48	1.73
N	Pink the Pig	0112	1.04	2.09	1.48
1	Bears	1111	1.25	1.77	1.49
I	Losing Teeth	1112	1.23	7.06	2.95
	Old Photos	2111	1.45	3.64	2.30
2	Game Inventor	2112	1.12	2.96	1.82
	Butterflies and Moths	2113	1.39	1.23	1.57
	Bad Talent Show	3111	1.14	1.98	1.50
3	Field Mice	3112	**	**	**
	Hamster on the Loose	3113	1.13	1.63	1.36

 Table 6.16. Odds Ratio Estimated from Logistic Regression across Passages and Grades (Winter 2017)

*Estimation is not converged.

**CI = confidence interval.

In Approach 2, both the WCPM and sentence reading fluency are treated as dichotomous variables for given cutoff values. Table 6.17 exhibits 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. In general, the results show that classification accuracy statistics are in the range from good to excellent for all grades except for Grade K.

Table 6.17. Classification Accuracy Statistics: Using Raw Score of Sentence Reading Fluency to Screen Passage Reader (Passage WCPM Cutoff Value = 30) (Winter 2017)

Grade	Passage Title	Passage Code	Cutoff Value of Screening Test	F PR [*]	FN R [*]	SEN [*]	SEP [*]	BR*	OCA [*]
к	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

Grade	Passage Title	Passage Code	Cutoff Value of Screening Test	FP R [*]	FN R [*]	SEN [*]	SEP [*]	BR [∗]	OCA [*]
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
	Bad Talent Show	3111	15	0.33	0.02	0.98	0.67	0.99	0.98
3	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.

6.8. Examining the Reliability of the LanguaMetrics Machine Scores

The LanguaMetrics engine provides scores at a lower cost compared to that of human scoring. Similar to the scores produced by human scorers, the reliability and validity evidence of scores produced by the scoring engine should be provided to ensure the proper use of the engine in the implementation of automated scoring of oral reading fluency. To example the reliability of the machine scores, the consistency and agreement between the LanguaMetrics machine scores and human scorers were examined using Winter 2017 data (i.e., Winter 2017 passage reading responses). In this study, each student read a given passage aloud, and all responses were recorded using the LanguaMetrics system. Each response was then scored by both a machine (LanguaMetrics system) and a human based on the number of words read correctly in a minute (WCPM). A total of 10 human scorers scored the passage reading responses.

High consistency and agreement between human and machine scores are the foundation to replace human scoring with machine scoring for evaluating oral reading fluency. Typical measures of rater agreement include correlation, kappa, weighted kappa, and exact and adjacent agreement. The study did not address the relationship between machine scoring and external criteria that is usually referred to as validity evidence, such as construct, criterion, and consequential (long-term predictive) related evidence.

6.8.1. Handscoring Process

Passage reading responses from the Winter 2017 administration were handscored by Strategic Measurement and Evaluation, Inc. (SME) as part of this score reliability study. Scorers were randomly assigned to serve as the first human rater or the second human rater for each response to check rater reliability.

6.8.1.1. Rangefinding

To meet professional testing standards, clear training materials need to exist that support the valid and reliable scoring of student responses. Scorers must understand what counts as an appropriate response and what does not and how to distinguish between the two. The goal of rangefinding was to develop a generalized set of training materials to identify key scoring issues to be used as a reference to score student responses. To begin this process, SME reviewed a sample of student responses to each of the 27 available passages (approximately 20 responses per passage). Two SME senior scoring leaders independently scored each oral reading response and recorded their scores on paper sheets (outside of the LanguaMetrics online system). After scoring, the senior scoring agreement and disagreement. This comparison also led to general discussions about scoring challenges experienced during scoring.

In reviewing areas of disagreement and challenges, the senior scorers identified the underlying issue that led to the disagreement (e.g., issues with phonemic awareness, students repeating words and moving from incorrect initial pronunciation to correct final pronunciation, misplaced emphasis, word substitutions, transposing word order, and lack of attention to punctuation when reading). Once the underlying issues were defined, the scorers could identify examples of each issue by referencing specific parts of sample student responses that illustrated the issue.

Based on these discussions, SME created a summary document that identified each issue and included references to specific examples of the issue (i.e., "Grade 2 passage 1; student 14; audio response time 0:20 to 0:25; word substitution"). This summary document was shared with NWEA content experts so that final scoring guidelines could be established. SME used the feedback provided by NWEA content experts to establish final scoring rules around each general oral reading fluency issue and finalized sample audio training sets with examples of acceptable and unacceptable responses. SME used this information to build annotated training materials to match the NWEA-approved audio samples and explain why each training sample either did or did not receive credit.

For example, during the scoring of the sample audio files provided by NWEA, senior scoring leaders noticed students sometimes mispronounced a word but then immediately corrected themselves and said the word correctly. While this situation was common and an appropriate scoring rule may seem obvious, a clear final scoring rule had to be established regarding when the initial vs. the corrected pronunciation would be the focus of scoring (e.g., "If a student immediately corrects a mispronunciation (less than 3 seconds), the word will be scored as correct").

The goal for the SME-developed training materials was to include examples of student responses across many passages. The focus of the materials was on the general higher-level concepts that shape scoring decisions. This type of approach to training is more efficient, memorable, and generalizable compared to trying to identify a large set of passage-specific samples for each passage. In addition to the training materials, SME built qualifying sets to be administered to potential scorers after the initial training. The qualifying sets included sample audio files associated with each of the general scoring issues identified in the first step and included in the training materials.

6.8.1.2. Scorer Training

SME was responsible for recruiting and training all scorers. All scorers had college degrees and were native English speakers. SME followed a general four-step process to score student responses for the MAP Reading Fluency assessment.

Step 1. The scoring leaders provided the scoring trainees with an overview of the project and showed them a sample passage that students were asked to read. They then played a sample audio response. This type of experience gave the trainees a chance to familiarize themselves with the content of the items they were asked to score, and it helped them build a deeper understanding of the skills needed to answer the item correctly.

- Step 2. The scoring leaders reviewed the scoring procedures and guidelines for this project and walked the trainees through the key features of each score point. As part of this discussion, the scoring leaders reviewed some commons issues and provided guidance on how the issues should be scored. The scoring leaders also reviewed the guidelines for assigning condition codes (e.g., insertions, interruptions, technical errors, unintelligible, incoherent, or sporadic speech, or off-topic speech) and provided instructions on how these should be scored.
- Step 3. A training set of five sample responses was presented, and the trainees assigned a score to each sample response. The trainees were presented with sets of complete one-minute audio clips that were selected by the senior scoring team to represent one or more coding or scoring-related issues. This training occurred outside of the LanguaMetrics platform. The lead trainer reviewed each sample response with the training group and facilitated a discussion. Trainees explained (using the words of the rubric) why they assigned the score they did. After discussing potential scores, the training leader identified the official score for the response and explained why it was the assigned score. The training sets included sample responses covering the range of possible score points for the item. The goal of this process was to calibrate all scorers to the same scoring rules so that the group shared a common understanding of how to apply the scoring rules.
- Step 4. Once the group completed the first training set, the trainees independently scored a final sample of five responses. Again, these training examples were full-length one-minute audio clips that were representative of the types of responses that trainees were asked to score. This second set of training responses was referred to as a qualification set because it was designed as a posttest to confirm that each scorer understood the scoring rules before moving forward to scoring live student responses. A passing score on the qualification set established that the scorer was qualified to score student work. Only trainees reaching the acceptance criteria established in collaboration with NWEA content experts (currently 4 of 5 scores matching the expert-assigned score) were approved for scoring the field test responses. For the Winter 2017 test event, all 10 scorers qualified.

6.8.2. Study Results

The following two methods, which are common in evaluating scoring consistency, were calculated for MAP Reading Fluency:

- 1. Inter-rater reliability (IRR): A correlation between two scorers' scores that measures the strength of a linear association of scores
- 2. Inter-rater agreement (IRA): Measures the extent to which two scores agree on the absolute value and measures agreement

Both methods examine how well the machine scores match the human scores, but they do have some distinctions. IRR measures consistency between two scorers in ordering or relative standing of scores without being concerned with absolute values of the scores. For example, correlations between two raters' scores can be perfect (i.e., 1), but the means of the two scores can be very different. IRR focuses on the consistency of raters' scores at the relative levels of performance, whereas IRA focuses on the consistency of raters' scores at the absolute level of performance.

The most commonly used method is for IRR is the Pearson product-moment correlation. For IRA, one of the following indices can be used:

- Percentage of exact agreement
- Cohen's kappa
- Intra-class correlation (ICC)

Because a rater's score of WCPM was an ordered integer for human and a continuous variable that can be rounded to an integer for machine, and both scores have large numbers of categories, it is not very meaningful to calculate kappa (as a chance-corrected measure of agreement) or the percentage of exact agreement under these circumstances. The ICC, on the other hand, can be used to estimate agreement based on a rating that is more than five categories or is a continuous variable (ICC can be used for ordinal, interval, and ratio variables). The advantage of ICC over the Pearson correlation is that the ICC is sensitive to both association and mean difference (bias) between scorers, while the Pearson correlation only measures association between scorers. The ICC has a range from 0 to 1 for two scorers and can be less than -1 for three or more scorers. An ICC of 1 represents a perfect agreement, and an ICC of 0 represents random agreement.

Table 6.18 presents the IRR and IRA results across passages and grades. The Pearson correlations are all higher than 0.80, suggesting that the human and machine scores are highly consistent across all grades and passages. The IRA results also demonstrate a strong agreement between the machine scores and human scores across all grades, as measured by the ICC.

Grade	Passage Title	Passage Code	Ν	IRR	IRA
K	Sal Gets Wet	0111	33	0.98	0.98
ĸ	Pink the Pig	0112	26	0.99	0.98
1	Bears	1111	152	0.87	0.85
1	Losing Teeth	1112	143	0.81	0.82
	Old Photos	2111	313	0.97	0.96
2	Game Inventor	2112	286	0.94	0.93
	Butterflies and Moths	2113	291	0.96	0.95
3	Bad Talent Show	3111	315	0.98	0.97
	Field Mice	3112	322	0.91	0.90
	Hamster on the Loose	3113	310	0.94	0.94

Table 6.18. IRR and IRA Score Reliability Results: Machine vs. Human Scores for PassageReading Responses (Winter 2017)

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge: MIT Press.
- 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.
- 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). New York: Guilford.
- Cassady, J. C., & Smith, L. L. (2004). Acquisition of blending skills: Comparisons among bodycoda, 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.
- 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. Washington, DC: CCSSO & National Governors Association.
- 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. Washington, DC: Government Printing Office.
- Deeney, T. A. (2010). One-minute fluency measures: Mixed messages in assessment and instruction. *The Reading Teacher*, *63*(6), 440–450.

- 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). Mahwah, NJ, US: 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*. London: 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., 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*. Houston, TX: 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). London: Lawrence Erlbaum Associates.
- 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.
- 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*. New York, NY: Guilford Press.
- Gillon, G. T. (2017). *Phonological awareness: From research to practice (second edition)*. New York, NY: Guilford Press.

- 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.
- Hasbrouck, J. & Tindal, G. (2017). *An update to compiled ORF norms* (technical report no. 1702). Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- 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.
- 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.
- 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.
- 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.
- Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational psychology*, *100*(2), 310.
- 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. Retrieved from <u>https://cdn.lexile.com/cms_page_media/135/The%20Lexile%20Framework%20for%20R</u> <u>eading.pdf</u>.
- Leslie, L., & Caldwell, J. (2006). Qualitative reading inventory 4. Boston, MA: 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.
- Lipson, M. Y., & Wixson, K. K. (2012). Assessment and instruction of reading and writing difficulties: An interactive approach. Pearson Higher Ed.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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). Boston, MA: Sopris West. Retrieved from <u>http://www.readingrockets.org/article/development-</u> <u>phonological-skills</u>.
- 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.
- 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.
- 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.
- Morris, D., Bloodgood, J. W., Perney, J., Frye, E. M., Kucan, L., Trathen, W., ... & Schlagal, R. (2011). Validating craft knowledge: An empirical examination of elementary-grade students' performance on an informal reading assessment. *The Elementary School Journal*, 112(2), 205–233.

- 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. Retrieved from <u>https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf</u>.
- 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). New York, NY: Guilford Press.
- Newman, H. M. (2009). *The influence of mandated oral reading fluency assessments on teachers' thinking and practice*. Unpublished doctoral dissertation, University of Washington, Seattle.
- Nitko, A. J. (2004). Educational assessments of students. Englewood Cliffs, NK: Prentice Hall.
- 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.
- O'Connor, R. E., Bell, K. M., Harty, K. R., Larkin, L. K., Sackor, S. M., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology*, *94*(3), 474.
- 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.
- Paulson, L. H. (2004). The development of phonological awareness skills in preschool children: From syllables to phonemes. Dissertation Abstracts International, University Microfilms No. 3166292. Retrieved from 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). Mahwah, NJ: 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.
- 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 Kamil, M. L., Pearson, P. D., Moje, E. B., & Afflerbach, P. (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 graphophonemic skills in elementary students. *Assessment for Effective Intervention*, *36*(1), 21–34.
- Samuels, S. J. (2006). Toward a model of reading fluency. In Samuels, S. J. & Farstrup, A. E. (Eds), *What research has to say about fluency instruction* (pp. 24–46). Newark, DE: 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). Timonium, MD: York Press.
- Schrank, F. A., Mather, N., & Woodcock, R. W. (2004). *Woodcock Johnson III Diagnostic Reading Battery: Comprehensive manual.* Itasca, IL: 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). New York, NY: 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.
- 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.
- 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.
- U.S. Census Bureau (2017). U.S. Census Bureau QuickFacts: Population estimates, July 1, 2017 (V2017). Retrieved from https://www.census.gov/quickfacts/fact/table/US/PST045217.