MOBY.READ

Technical Manual

November 2018

Analytic Measures Inc.

Contents

List of Figures

List of Tables

1. Introduction

Moby.Read is a self-administered oral reading fluency (ORF) assessment that runs on iOS and HTML5-compliant devices and is automatically scored. The test measures oral reading fluency in English of students in grades 1 through 5. Students read a word list, a sentence, and four short passages out loud, and they are prompted to retell the passages and respond to short-answer questions using their own voice. Moby.Read uses automatic speech recognition (ASR) and natural language processing (NLP) technologies to score oral reading fluency and comprehension. An integrated cloud-based administrator interface offers aggregated performance charts and access to student recordings, which provides teachers with immediate, accurate, personalized data that support individual instruction to improve student achievement.

The *Moby.Read Technical Manual* is intended for educational professionals, such as administrators, teachers, or reading specialists, who are responsible for evaluating the oral reading ability of students in grades 1 through 5.

2. Test Construct

The Moby.Read assessment measures oral reading fluency (ORF) in English of students in grades 1 through 5.

ORF is defined as "learning to recognize (decode) words in a passage automatically (effortlessly) as well as accurately and to express or interpret those words in a meaningful manner when reading orally" (Rasinski, Padak, McKeon, Wilfong, Friedauer & Heim, 2005).

Research has identified four components of oral reading fluency: accuracy, rate, prosody (or expression), and comprehension (Deeney, 2010). These core competencies align with key aspects of reading fluency supported by educational standards. For example, the Common Core State Standards (NGA & CCSSO, 2010) asserts in its Foundational Skills in the English Language Arts Standards that students should be able to:

> Read with sufficient accuracy and fluency to support *comprehension*. (CCSS.ELA-LITERACY.RF.2.4)
>
> Read grade-level text orally with *accuracy*, appropriate *rate*, and *expression* on successive readings. (CCSS.ELA-LITERACY.RF.X.4.B)

Even in states that have not adopted CCSS, such as Texas, for example, state educational standards include reading fluency as a part of foundational language skills, emphasizing the same four core competencies:

> The student reads grade-level text with fluency and *comprehension*. The student is expected to use appropriate fluency (*rate, accuracy,* and *prosody*) when reading grade-level text. (TEA, 2017)

The Moby.Read assessment measures all four ORF components—Comprehension, Accuracy, Accurate Reading Rate, and Expression—on grade-Leveled texts for students in grades 1 through 5. In addition, the test reports an overall Moby.Read Level that integrates all four components into a single measure.

3. Intended Use of the Assessment

Moby.Read is designed as an oral reading fluency benchmark assessment that provides reliable results of key fluency measures. Test scores can be used for a variety of administrative and pedagogical purposes. Moby.Read Levels can help determine benchmark levels of oral reading

performance specific for a student's grade and time of year. Administrators can use test results for state-level reporting purposes. Teachers can use Moby.Read Level scores for personalizing instruction and choosing appropriate reading assignments; they can use Comprehension scores to evaluate whether a student needs help decoding and constructing meaning from texts; teachers can analyze Accuracy scores to understand whether a student needs practice in word recognition; Rate scores may indicate that a student needs support with automaticity; and Expression scores can help identify students who may need more practice capturing the meaning of passages through phrasing.

Moby.Read scores are color-coded to facilitate interpretation and subsequent grouping of students with similar scores for tailored instructional intervention. For example, a student performing at a below-target Moby.Read Level and Accuracy score (color-coded red) can be placed in a group of students with similar scores, and the teacher can work on phonics instruction with this group. A student underperforming in both Comprehension and Moby.Read Level might benefit from creating content maps after a reading assignment. Pairing students with complementary strengths and weaknesses and having them work together as a team can motivate students to perform better. To improve Expression scores, a teacher might organize a reader's theater project focusing on expressive reading. These examples illustrate how Moby.Read scores might be used in a classroom setting.

3.1 Appropriate Uses

When evaluating students' oral reading performance, Moby.Read Levels should be considered in the context of all available information about a student and with professional input from teachers and reading specialists. Although Moby.Read Levels may help guide the selection of reading assignments, Level scores should never dictate text difficulty in isolation. Research suggests that with appropriate instruction, students who practice reading at levels higher than their assessed level can show considerable gains in performance (Brown, Mohr, Wilcox, & Barrett, 2017; Morgan, Wilcox, & Eldredge, 2000). For others, motivation may be a concern, and high-interest texts below their reported level may be appropriate. In all cases, Moby.Read Levels are only one piece of information to consider when tailoring instruction to the needs of individual students.

Any instructional intervention based on Moby.Read Accurate Reading Rate scores should address the broad construct of oral reading fluency. Some instructors view the achievement of benchmark accuracy rates as the end-goal, placing undue emphasis on encouraging students to read fast at the expense of reading for meaning with appropriate expression (Deeney, 2010). All instruction should emphasize a broad construct of fluency, helping students exercise all core competencies of oral reading fluency—including Comprehension and Expression—instead of focusing only on Accuracy and Rate. Practice will naturally increase a student's reading rate (Herman, 1985; Therrien, 2004).

Moby.Read scores indicating high performance should never justify lack of instruction. Even students excelling at oral reading fluency should be encouraged to develop their reading skills further and all students should be broadly supported in their efforts. Students at any Level can improve their reading skills with appropriate instruction and practice.

4. Test Description

The Moby.Read assessment consists of several sections as shown in Figure 1. The session begins with a video introduction showing a student taking the test and reading out loud. The video introduces the tasks and provides a model for students, reminding them to read for meaning and to read out loud with confidence.

Following the video, students are prompted to read a word list and a sentence out loud. The remainder of the test consists of four passages. The first one is a practice passage that is set at a grade level below the other operational passages and is not scored. Only the three operational passages are scored.

Students are asked to read each passage, retell what they have read in their own words, and then answer two questions about the passage content. Students who wish to read the passage again may select that option and hear a fluent reading of the passage with synchronized text highlighting. The optional reread is not scored.

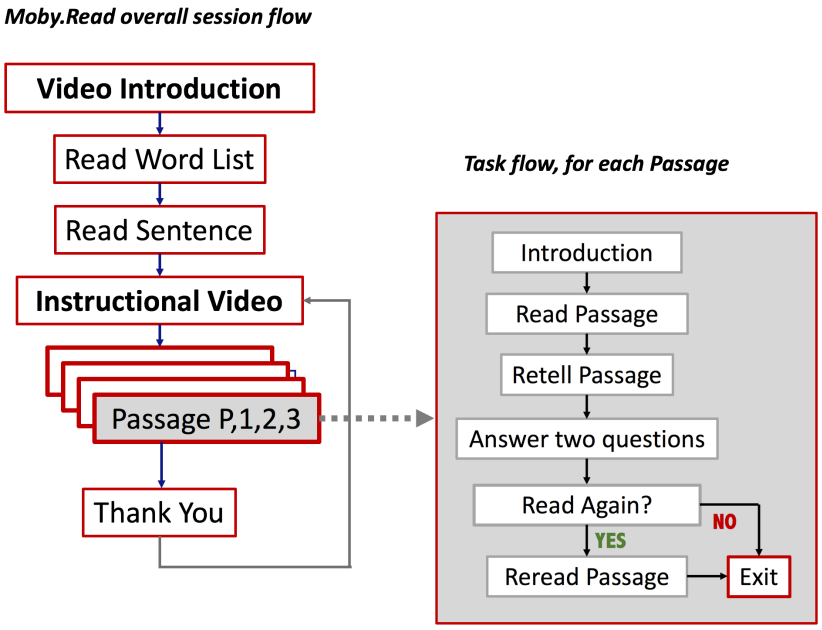

Figure 1: Moby.Read session flow

The Moby.Read test is self-administered by students. This means that once the test starts, no teacher or administrator intervention is required to move the test along. The application detects the student's speech while the student is responding to prompts and advances to the next task after a set amount of time. Remaining time is shown in a yellow progress bar on top of the screen. A student who finishes a task early can advance the test by pressing the "Next" button (active only after the student has been reading for a set amount of time). If a student has not finished the task before the system's automated response time-out, the test advances to the next task.

For most students, a Moby.Read assessment takes approximately 12 minutes per test form.

4.1 Test Structure

The overall test structure is summarized in Table 1.

Table 1: Moby.Read test sections

Section	Presented for Each Administration?
Introductory Section	
Video Introduction	No, only first time per user performance
Read Word List	No, only first time per user performance
Read Sentence	No, only first time per user performance
Practice Passage (not scored)	
Model Reader Video	Yes
Read Passage P (Practice)	Yes
Retell Passage P (Practice)	Yes
Short-Answer Question 1 for Practice Passage P	Yes
Short-Answer Question 2 for Practice Passage P	Yes
Model Reading and Reread	Optional
Scored Passages 1 - 3	
Read Passage	Yes
Retell Passage	Yes
Short-Answer Question 1	Yes
Short-Answer Question 2	Yes
Model Reading and Reread	Optional (not scored)

4.1.1 4.1.1 Introductory Section

The introductory section is presented only during the first time a student ID is used for a given test form. It includes an introductory video, a word list, and a sentence read. The video introduces the types of tasks presented in the test and shows a student performing the tasks. The narrator reminds students to read for meaning and to speak in a clear and strong voice when reading out loud or answering questions. Following the video, students are asked to read a list of words and a single sentence out loud.

4.1.2 Practice Passage

A practice passage introduces the passage reading tasks. The practice passage is not scored and is leveled lower than the grade-leveled passages. The practice passage is meant to familiarize students with the technology and format of the assessment.

Model Reader video. Like the three scored passages, the practice session starts with a video explaining the tasks. The video shows a student confidently reading a passage out loud, retelling the passage in his own words, and answering two questions about the passage.

Passage task instructions. The test introduces the practice passage by displaying the title and a picture on the screen while a narrator explains the task. The display changes to showing the passage text underneath the picture. The narrator instructs the student to “start reading here,” while the first few characters flash in red. The narrator reminds students to remember what they read and invites them to retell the passage in their own words.

Read passage. Students read the passage displayed on the screen out loud.

Retell passage. After reading the passage, students are asked to retell what they have read in their own words.

Answer short-answer questions. After completing the passage retelling task, students are prompted to answer two short-answer questions about the passage they just read. Each question is read aloud by the narrator and displayed in writing on the screen. The student answers each question aloud.

Listen to model reading and reread passage. After retelling the passage and answering questions, students can listen to the narrator reading the passage and then read it out loud again themselves. The rereading task is optional; it provides an opportunity to hear a model reading and additional reading practice, but responses are not scored.

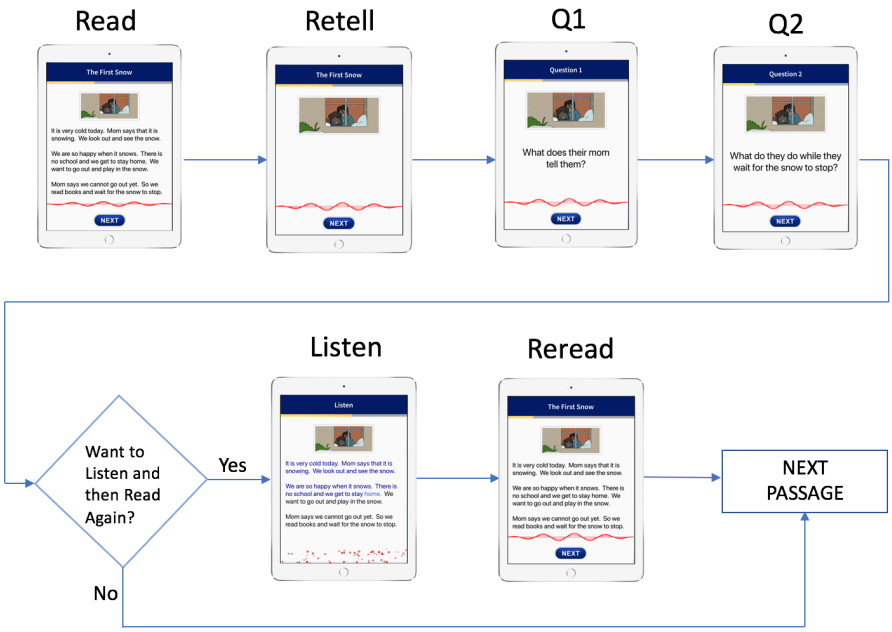

Figure 2: Moby.Read passage block flow

4.2 Scored Passages

Three scored passages follow the practice passage. A narrated introduction is followed by a passage reading, a passage retelling, two short-answer questions, and an optional task of listening to a model reading and rereading the passage. The optional rereading is for practice only and is not scored. Figure 2 shows the task flow for a single test session. .

Scored passages are presented at grade level. Moby.Read test forms typically include two narrative passages and one informational passage, with the easiest narrative passage presented first and the informational passage presented second. In some 4th and 5th grade forms, a second informational passage may be used to optimize scoring for empirical difficulty of the form. The text types (narrative, informational) per form are presented in the Appendix, Moby.Read Text Types.

5. Content Development

A team of education specialists, linguists and reading professionals with advanced degrees and experience designing reading assessments developed the items included in the test. All items were reviewed for bias and fairness and were tested in extensive field trials involving 982 students in schools across four states.

5.1 Task Design and Material Selection

Moby.Read items and related tasks were designed to elicit responses that cover all relevant grade-level ORF competencies and to provide sufficient speech data to ensure reliable machine scoring.

5.1.1 Word Lists and Sentence Reads

Word lists and sentence reads were selected to elicit responses requiring lexical and semantic decoding abilities relevant to the grade level.

5.1.2 Passage Tasks

Spoken instructions introduce each passage providing context for the passage and introducing key vocabulary that students may encounter in the passage.

Passages are one of two content types: (1) literary texts, such as short narratives or stories and (2) informational texts covering topics such as history, art, science, or technology. These text types align with common content types found in most national assessments and state reading standards (NAEP 2017, NGA & CCSSO, 2010). To ensure appropriate domain sampling, each form of the Moby.Read test includes both narrative and informational texts.

Passages are roughly 40 to 150 words in length and are designed to provide enough spoken material for reliable machine scoring while being short enough to allow for efficient testing and timely test completion by a typical student of a given grade level (Hasbrouck and Tindal, 2006).

All passages begin with a short lead-in sentence that uses familiar words and uncomplicated spelling. The simplicity of the lead-in sentence eases students into reading with confidence. Sentence complexity and vocabulary difficulty increase as the passage progresses.

5.1.3 Spoken Retellings

The passage retelling task is included for two reasons. First, retelling a story requires that the reader is reading for meaning and has understood the passage content. The extent to which retelling captures key elements of the reading is a measure of the reader's comprehension. Second, retelling is an established strategy for encouraging deeper text processing (Morrow, 1985; Wilson, Gambrel & Pfeiffer, 1985).

Analytic Measures Inc. (AMI) used a three-phased approach to passage levelling, that combines quantitative and qualitative analysis with empirical data from field trials. See section 5.3.3.

5.1.4 Short-Answer Questions

Short-answer questions are designed to elicit a single word, or a short phrase or sentence in response. Questions vary in difficulty and degree of cognitive processing required, ranging from literal knowledge-based questions to more complex questions that require deeper, inferential thinking. Most Moby.Read short-answer questions are direct and literal at the early elementary grades, with more inference-based questions added in grades 4-5.

5.2 Review for Bias and Fairness

All content was reviewed for bias and for any content that may negatively affect reader performance. Where appropriate, items were revised or underwent further vetting. A team of experts with advanced degrees in psychology and psycholinguistics and experience in reading assessment reviewed the items. Additionally, an independent review was conducted in January 2018 by Educational Testing Service (ETS) to ensure that all Moby.Read materials were aligned with ETS's *Guidelines for Fair Tests and Communications* (ETS, 2015). Another ETS review was conducted in July 2018 reflecting updates to the guidelines. In both cases, passages or prompts that did not meet the guidelines were either not included in the test or were revised to align with ETS guidelines.

5.3 Field Testing

All Moby.Read items were piloted with students in several field trials to ensure they were working as intended. In addition, field testing served to collect speech data for training the speech recognition engine and developing machine scoring algorithms. Usability data was collected through a brief survey asking participants to rate their experience with the Moby.Read app. Field trials were conducted from October 2016 through May 2018.

5.3.1 Field Test Participants

A total of 982 students took part in the Moby.Read pilot. Table 2 through Table 4 show demographic information where available. Many schools chose to withhold demographic information to protect student privacy.

Table 2: Participant state of residence

State	N	%
California	274	27.9%
Colorado	228	23.2%
New Jersey	147	15.0%
Ohio	333	33.9%
TOTAL	982	100%

Table 3: Participant gender

Gender	N	%
Female	289	29.4%
Male	371	37.8%
[unknown]	322	32.8%
TOTAL	982	100%

Table 4: Participant grade

Grade	N	%
1st	267	27.2%
2nd	173	17.6%
3rd	189	19.2%
4th	110	11.2%
5th	44	4.5%
6th	61	6.2%
7th	3	0.3%
8th	1	0.1%
9th	1	0.1%
[unknown]	133	13.5%
TOTAL	982	100%

5.3.2 Field Test Protocol

Participants took the Moby.Read field test on a range of platforms, including iPads, Chromebooks, and laptops with HTML5-compliant web browsers. When available, microphone headsets were used. Tests were self-administered, advancing through tasks based on user input or on the system's automated response time-out. A human proctor monitored several concurrent sessions.

Several test-form variants were used with separate groups of participants. For some test administrations, additional word lists were presented before the passages. Most forms included four passages (one practice passage and three scored passages). In some cases, additional passages were included depending on students' reading ability, observed reading endurance, and interest. In these cases, after the student completed a test with three or four passages, the administrator manually started an additional test, which was then automatically administered. Some trials were administered in close succession with other reading assessments such as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test.

5.3.3 Item Characteristics

A total of 82 passages were tested in the Moby.Read field trials. Passages were selected for inclusion in forms based on consistency between an *a priori* estimate of each passage's difficulty and empirical data. AMI used a three-phased approach to estimate *a priori* passage difficulty, which aligns with similar approaches proposed by current standards (CCSSO & NGA, 2012). First, a quantitative levelling was performed on all passages. The Flesch-Kincaid text leveling method (Kincaid, Fishburne, Rogers & Chissom, 1975) was selected for quantitative leveling, given its broad use, longevity, and conceptually intuitive nature. Second, a qualitative analysis of each passage by AMI's expert team considered several textual dimensions: structure, language conventionality, clarity, knowledge demands, and levels of meaning (literary) or purpose (informational). When the quantitative text-complexity value differed significantly from the qualitative expert judgments, passages were set aside. In the final phase, human experts judged the appropriateness of the qualified passages for target readers performing the target task.

The resulting difficulty estimates were projected on a scale from mA (Moby A) to mZ (Moby Z) with three Levels per grade, except for the lower end of the scale. Table 5 presents the Moby.Read scale.

Table 5: Moby.Read scale for estimating passage difficulty

Moby.Read Level	Grade
mA	K (spring)
mB	
mC	1st (fall)
mD	
mE	1st (winter)
mF	
mG	1st (spring)
mH	
mI	2nd (fall)
mJ	2nd (winter)
mK	2nd (spring)
mL	3rd (fall)
mM	3rd (winter)
mN	3rd (spring)
mO	4th (fall)
mP	4th (winter)
mQ	4th (spring)
mR	5th (fall)
mS	5th (winter)
mT	5th (spring)
mU	6th (fall)
mV	6th (winter)
mW	6th (spring)
mX	above 6th
mY	above 6th
mZ	above 6th

Field test data were analyzed to determine how well students' actual reading performances matched estimated difficulty levels. AMI used a statistical model that determines a passage's difficulty level based on readers' observed Accurate Reading Rates with reference to the passage difficulty. The same model was used for scoring student Moby.Read Levels (see section 6.3.1 Moby.Read Level).

Passages that were most consistent with the *a priori* subjective estimates of passage difficulty were included in fixed forms for fall, winter, and spring benchmarks at each grade level.

Table 6 shows the empirical difficulty level of passages, averaged by grade and expressed in logits. This value is derived from a statistical model that was fitted to student data on these passages. A larger number of logits indicates a higher level of difficulty. The right column in Table 6 displays the average Moby.Read passage difficulty per grade, with difficulty defined as text complexity calculated according to Flesch-Kincaid.

Table 6: Average Moby.Read passage difficulty

Grade	Average Empirical Difficulty in Logits (Model from Accurate Reading Rate)	Average Text Complexity (Flesch-Kincaid)
1st Grade	-2.01	0.1
2nd Grade	-1.23	3.0
3rd Grade	0.15	3.9
4th Grade	1.81	5.4
5th Grade	3.39	6.1

As shown in Table 6, empirical difficulty gradually increases from grade to grade. Flesch-Kincaid values fall within the ranges recommended by the CCSS (CCSSO & NGA, 2012).

6. Scoring

Tests are automatically scored using automatic speech recognition (ASR) and natural language processing (NLP) technologies.

6.1 Speech Recognition System

The Moby.Read automatic speech recognition (ASR) system processes spoken responses in two steps. Spoken responses are first analyzed by the ASR system, and scoring algorithms are then applied to the ASR results to produce oral reading scores.

The ASR system deploys acoustic models and language models to analyze students' oral performance. The acoustic model used in the Moby.Read ASR system is a Deep Neural Network–Hidden Markov Model (DNN-HMM) (Zhang, Trmal, Povey, & Khudanpur, 2014) with four hidden layers. Language models for students' reading responses are rule-based and specific to each reading passage. Language models for retelling responses and short-answer questions are also item-specific. Given that the language models are item-specific and that correct readings are more likely than incorrect readings, the ASR system gives the reader the benefit of the doubt and generally assumes that a student has read a word correctly, even if the student reads with a dialectal form or a non-native accent. The Moby.Read ASR engine is based on Kaldi (Povey et al., 2011). The engine has been optimized for children's speech. Jurafsky & Martin's (2009) *Speech and Language Processing* covers most of the underlying machine learning methods discussed below.

In a second step, the time-aligned response strings generated by the ASR system are scored using certain specified criteria. For example, timing information at the syllable, word, and phrase level, along with inter-word silences, are used in Expression scoring.

6.2 Human Ratings

Development of scoring algorithms for Comprehension and Expression scores relied on human ratings. External raters started as trainees and were presented with rubrics (see sections 6.3.2 Comprehension and 6.3.5 Expression). Trainees were given training sets of actual student recordings to practice rating. These training ratings were then compared to expert ratings that had been agreed upon by a team of assessment professionals. Discrepancies between trainee ratings and expert ratings were reviewed, and additional training responses were presented to trainees until they were reliably assigning ratings consistent with the expert ratings. Only those

trainees whose ratings converged with AMI's experts' ratings were engaged as external raters and had their data used in the Moby.Read development process.

6.3 Score Components

The Moby.Read score report includes the following five score components:

1. Moby.Read Level (mA through mZ)
2. Comprehension Score (0 through 8)
3. Accuracy (percent)
4. Accurate Reading Rate (words correct per minute)
5. Expression (0 through 4)

6.3.1 Moby.Read Level

The Moby.Read Level is a single value representing the reader's current oral reading fluency ability. Levels are reported on an alphabetic scale from mA (Moby.Read Level A) through mZ (Moby.Read Level Z). Moby.Read Levels are reported on the same scale as passage difficulty presented in Table 5 and are mapped to grade levels in the same way. On this scale, there are three Moby.Read Levels for grades 2 through 6 and six Levels for grade 1.

Moby.Read Levels are based on a polytomous Rasch model. Using a joint maximum likelihood estimation technique, reader ability and passage difficulty are estimated concurrently on a logit scale. This scale is then mapped to Moby.Read Levels using an equation that links students' ability (based on Accurate Reading Rate ranges) and expectations of where students should be in their grade based on accurate reading rate norms published by Hasbrouck and Tindal (2006).

Because Accuracy and Comprehension are critical parts of oral reading fluency, adjustments to Moby.Read Levels based on Accuracy and Comprehension scores were included in the scoring algorithm. Expression is also critical but tends to covary with Rate and did not decrease or increase the Moby.Read Level.

If the reader's Rate was at the median or above grade-level norms but Accuracy was not at least 90%, the Moby.Read Level reflected a lower Level. This threshold of 90% was set lower (82%) for first graders based on empirical analysis.

To determine Level adjustments for Comprehension, human Comprehension scores for all available data were analyzed, and median Comprehension scores were derived for each Level. Comprehension scores two points below the median for a given Level triggered an exception in which no Moby.Read Level was assigned because of a low Comprehension score. See section 6.4.6 for more information on exception scoring.

Adjustments to Levels were also made at the higher end of the Moby.Read Level scale. Although Accurate Reading Rate is a robust predictor of reading ability, median Rate values taper off in middle school and remain stable throughout adulthood. Thus, reading faster beyond the median value does not necessarily indicate improved ability. The scoring algorithm adjusts Moby.Read Levels at Levels mU and higher based on Comprehension scores. Beyond this threshold, higher Moby.Read Levels are more associated with higher Comprehension scores than with higher Accurate Reading Rate scores.

6.3.2 Comprehension

Comprehension is a measure of the student's understanding of the material read. It indicates the degree to which a reader can identify or present the major and minor concepts, themes, and facts contained in a passage. Moby.Read Comprehension scores are reported on a scale from

0 to 8, with higher values representing greater Comprehension. Moby.Read Comprehension scores combine scores of retellings and of short-answer questions across passages. The reported Comprehension score gives equal weight to students' retelling of passages and to the two responses to short-answer questions, averaged across three scored passages.

6.3.2.1 Scoring Retellings

Human expert judgments of retelling responses were used to model the similarity between the meaning of a passage and a student's retelling. Human expert judgments of passage retelling responses were used to validate the machine-trained scoring model. An interdisciplinary team of nine human expert judges rated retelling responses on a six-point scale as shown in Table 7. Ratings from at least two human raters were collected for each retelling response and for each short-answer question.

To predict how a human would rate the retelling responses, a metric was created that considered equally (1) a quantification of the different number of word types common between the passage and the retelling response and (2) how semantically similar a spoken retelling is to the passage. The scoring algorithm for retelling responses measures the semantic similarity between the passage text and the retelling response using trained networks and natural language analysis. Specifically, Moby.Read uses Google's *word2vec* word vectors that were trained on about 100 billion words. The scoring model holds 300-dimensional vectors for 3 million words and phrases. More technical information about the applied technique is available in Cheng (2018). This metric was mapped on a 0 to 4 scale for the retelling component of the Comprehension score.

Table 7: Rubric for rating responses to passage retellings

Rating	Description
0	NOT RATED. Silent, irrelevant, or unintelligible.
1	MINIMAL.
2	LIMITED. Some concept sequences; missing major concepts and main narrative arc.
3	PARTIAL. Several concept sequences and related parts of the main narrative.
4	ADEQUATE. Enough major and minor concepts suggest main narrative logic.
5	GOOD. Major and minor concepts convey main narrative path and causal logic.
6	COMPLETE. All major and many minor concepts support close narrative fidelity.

6.3.2.2 Scoring Short-Answer Questions

For short-answer questions, human ratings informed the model for scoring student responses. Expert human judges also validated the machine scoring model for short-answer questions on a disjoint set of responses. A team of five raters with advanced degrees in psychology, psycholinguistics, and biology participated in rating short-answer questions. Responses were rated on a scale from 0 to 2 as shown in Table 8.

Table 8: Rubric for rating responses to short-answer questions

Rating	Description
0	Silent, irrelevant, or incorrect
1	Partially correct
2	Completely correct

The short-answer questions scores were each mapped to a 0 to 2 scale and then added together. The summed short-answer question score was then combined with the scores

derived from the retelling responses (mapped into a 0-4 range) and reported in an overall Comprehension score.

6.3.3 Accuracy

The Accuracy score is reported as the percentage of words read correctly over the number of words attempted. For Accuracy scoring, the word string produced by the ASR system is aligned with the passage text to determine the first and last word attempted by the reader. The speech-to-text alignment combined with the language model for the specific item type allowed for the development of an algorithm that calculates the percentage of words the student read correctly. Self-corrections are counted as correct, but omissions, substitutions, and severe mispronunciations are counted as incorrect.

6.3.4 Accurate Reading Rate

Accurate Reading Rate is the rate at which a student is reading words correctly. Accurate Reading Rate is reported as Words Correct Per Minute (WCPM) and typically ranges from 0 WCPM to 250 WCPM. Even if the student 's overall reading time is less than a minute, the Rate is computed as WCPM. For each scored passage, the words read correctly between the first and last text-relevant word attempted are added up. The total is divided by the time between the onset of the first word attempted and the offset of the last word attempted. The median WCPM across the three scored passages is the reported WCPM score.

6.3.5 Expression

Expression is the degree to which a student can express meaning and text structure when reading a passage aloud using appropriate rhythm, phrasing, intonation, and emphasis. Expressive reading enhances the listener's understanding and enjoyment of a text (Miller & Schwanenflugel, 2008).

Scoring models for Expression were based on human ratings from experts and from skilled raters trained by AMI's experts to produce accurate ratings of Expression. Training data for the machine-based Expression-scoring models included at least two human ratings of Expression for each passage reading. Human raters used a 6-point scale to rate passage readings, as shown in Table 9.

Table 9: Rubric for rating Expression

Rating	Description
0	Insufficient sample for rating.
1	Word-by-word rendition with no reflection of word, phrase or sentence meaning.
2	Some local word grouping; little sentential phrasing.
3	Exhibits some text-inappropriate phrasing; sentence- and passage-level meaning is partially conveyed.
4	Prosody generally reflects meaning, but phrasing or intonation is sometimes inconsistent with the text.
5	Read for a listener; intonation, phrasing, and emphasis appropriately express the meaning of the passage.

Human ratings were used to train a neural network with the goal of predicting how a human rater would rate the Expression of a passage reading. The features used in the neural network were produced by the ASR system and included the pattern of phonetic segment durations and

the log likelihoods of inter-word silence durations. The output values from the neural network were then mapped to a 5-level Expression scale from 0 to 4.

6.4 Standards and Cut Scores

Scores reported are color-coded to facilitate easy interpretation. Color-coded longitudinal performance graphs provide a quick, intuitive view of students' progress as shown in Figure 3.

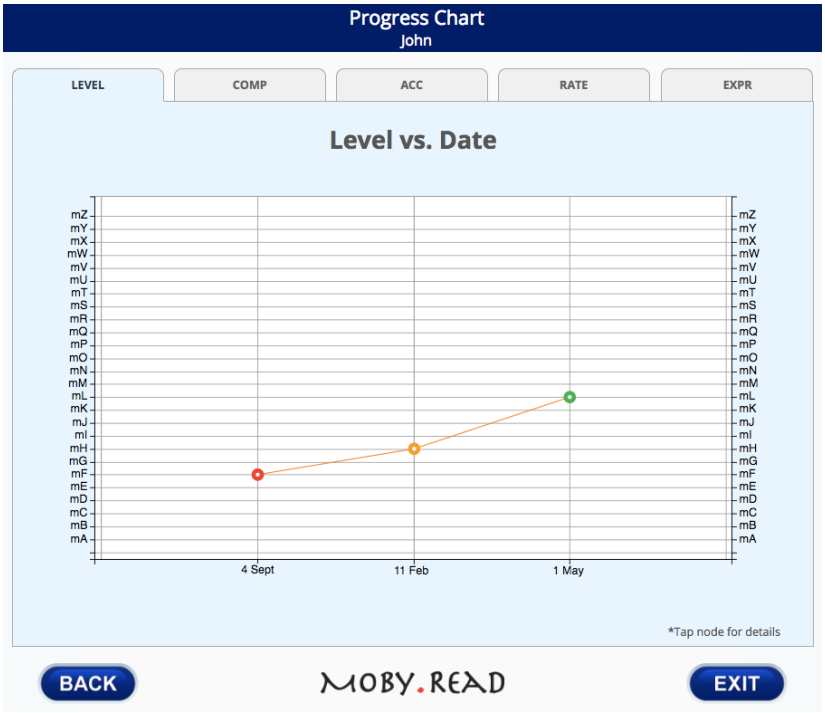

Figure 3: Color-coded student progress chart

Scores are coded as follows: Green indicates scores at or above grade level; orange-colored scores are slightly below grade-level; and below-grade-level scores are flagged as red. AMI used different methods for determining cut scores and defining the color coding for each Moby.Read score.

6.4.1 Cut Scores for Moby.Read Levels

Student grade and time of year of the assessment are both factors in considering whether a student's Moby.Read Level is below grade level. The Moby.Read Level scale includes at least three Levels per grade, typically associated with time of year, fall, winter, and spring.

A team of three reading assessment professionals defined Moby.Read Level cut scores as follows: If a student scores two or three Levels below the target Moby.Read Level for the student's grade and time of year, performance is below target (orange). A student scoring four or more Moby.Read Levels below the target is considered at-risk (red) and in need of intervention. At-risk performance is assumed to be more than one grade Level below the student's current grade. All other Levels are presented in green. Table 10 presents Moby.Read Levels for three performance levels.

Table 10: Ranges of Moby.Read Levels for three performance levels

Grade	Season	At Risk (Red)	Below Level (Orange)	At Level (Green)
1	Fall	—	—	mA+
	Winter	—	mA-mB	mC+
	Spring	mA	mB-mD	mE+
2	Fall	mA-mE	mF-mG	mH+
	Winter	mA-mF	mG-mH	mI+
	Spring	mA-mG	mH-mI	mJ+
3	Fall	mA-mH	mI-mJ	mK+
	Winter	mA-mI	mJ-mK	mL+
	Spring	mA-mJ	mK-mL	mM+
4	Fall	mA-mK	mL-mM	mN+
	Winter	mA-mL	mM-mN	mO+
	Spring	mA-mM	mN-mO	mP+
5	Fall	mA-mN	mO-mP	mQ+
	Winter	mA-mO	mP-mQ	mR+
	Spring	mA-mP	mQ-mR	mS+

6.4.2 Cut Scores for Comprehension

Standards for Comprehension scores are criterion-based. Scores were adjusted for developmental appropriateness, since retelling a passage is a cognitively demanding task, especially for younger students. Across all grades, a Comprehension score of 4.0 or more was considered at grade level since this level is approaching satisfactory Comprehension.

To determine what was developmentally appropriate, Comprehension scores from all available data were analyzed and median Comprehension scores were calculated for each grade level. Cut scores for this data set were defined by a team of three reading assessment professionals. The cut scores were determined to be 1.5 points below the median for below-level targets (orange) and 2.5 points below the median for at-risk targets (red). As a result, a Comprehension score of 2 is marked in red as below-grade level for higher grades (4 and 5) but not for lower grades (1 and 2). Table 11 shows score ranges per performance levels for grades 1 through 5.

Table 11: Moby.Read Comprehension score ranges for three performance levels

Grade	At Risk (Red)	Below Level (Orange)	At Level (Green)
1	0 – 0.5	1.0-1.5	2.0+
2	0 – 0.5	1.0-1.5	2.0+
3	0-1.5	2.0-2.5	3.0+
4	0-2.5	3.0-3.5	4.0+
5	0-2.5	3.0-3.5	4.0+

6.4.3 Cut Scores for Accuracy

Standards for Accuracy percentages were adopted from reading pedagogy. When a student reads less than 90% of a text accurately, the text may cause frustration, thus below 90% Accuracy is sometimes known as the frustration level (Johnson & Kress, 1965). Therefore, Moby.Read flags Accuracy scores below 90% as red. Values above the 90% threshold are

taken to be within the student's grade-level reading ability (marked as green). Percentage score ranges associated with performance levels are shown in Table 12.

Table 12: Moby.Read Accuracy score percentage ranges for two performance levels

Grade	At Risk (Red)	At Level (Green)
1-5	0%-89%	90%+

6.4.4 Cut Scores for Expression

Standards for Moby.Read Expression scores are criterion-based and were determined by a team of three reading assessment professionals. Scores of 3 and 4 (mostly fluent and fully fluent) were considered at-level performance (green). Scores of 2 (partially fluent) were classified as below-target performance (orange). Scores of 0 and 1 (no evidence of Expression and word-by-word reading) were designated as at-risk performance (red). Table 13 summarizes the Expression score ranges associated with three performance levels.

Table 13: Moby.Read Expression score ranges for three performance levels

Grade	At Risk (Red)	Below Level (Orange)	At Level (Green)
1-5	0-1	2	3-4

6.4.5 Cut Scores for Accurate Reading Rate

To identify cut scores for Accurate Reading Rate (WCPM), AMI referenced published norms from Hasbrouck and Tindal (2006). The same group of assessment professionals that provided the models for human expert ratings determined that 10 to 19 words below the median WCPM for a given time of year per grade was considered lower than grade-level target performance (orange), and scores of 20 or more below the median were considered at-risk (red). Based on the observation that the median of Accurate Reading Rate typically increases by 20 words per grade, the experts concluded that scores of 20 or more below the median reflect a performance one full grade level below the current grade. Table 14 presents Moby.Read Accurate Reading Rate ranges associated with three performance levels. Median Rate values for fall, winter and spring are taken from Hasbrouck & Tindal (2006).

Table 14: Moby.Read Accurate Reading Rate score ranges for three performance levels.

Grade	Season	At Risk (Red)	Below Level (Orange)	At Level (Green)	Median Rate
1	Winter	0-3	4-13	14+	23
	Spring	0-33	34-43	44+	53
2	Fall	0-31	32-41	42+	51
	Winter	0-52	53-62	63+	72
	Spring	0-69	70-79	80+	89
3	Fall	0-51	52-61	62+	71
	Winter	0-72	73-82	83+	92
	Spring	0-87	88-97	98+	107
4	Fall	0-74	75-84	85+	94
	Winter	0-92	93-102	103+	112
	Spring	0-103	104-113	114+	123
5	Fall	0-90	91-100	101+	110
	Winter	0-107	108-117	118+	127
	Spring	0-119	120-129	130+	139

6.4.6 Exception Scoring

Among scored passages, any of the following events triggers exception scoring:

1. Speech time is less than ten seconds
2. Less than ten correct words are recognized
3. Words Correct Per Minute (WCPM) is below ten or above 300

If one of three scored passages triggers an exception, the WCPM score is calculated as the average of the remaining two passages. Accuracy, Comprehension, and Moby.Read Level are similarly calculated from the remaining passages alone. If the student produces more than twelve correct words in the remaining passages, Expression is calculated as the average of those passages; otherwise, an Expression score of zero is returned.

If two or more scored passages trigger an exception in any given test session, all scores are reported as zero, and no Moby.Read Level is reported.

Comprehension scores two points below the median score associated with a student's calculated Moby.Read Level are reported as an exception in which no Level is returned.

The Moby.Read score report marks exceptions as a black X icon:  .

7. Validity Evidence

Validity evidence substantiates that a test is measuring what it purports to measure and thereby supports specific uses of test scores. Sources of validity evidence include expert verification as well as empirical data.

7.1 Evidence of Content Validity

Content validity is the extent to which items on a test cover a representative sample of material appropriate for the test domain. For the Moby.Read test, the domain is *written text* appropriate for students in grades 1 through 5. Moby.Read defines the appropriate domain in terms of two content types, narratives and informational texts, in accordance with national assessments and common standards of reading (NAEP 2017, NGA & CCSSO, 2010) and includes both content types to ensure appropriate domain sampling. By including a balance of carefully leveled narrative and informational texts, the Moby.Read test content is representative of the broad domain of texts appropriate for grades 1 through 5.

7.2 Evidence of Construct Validity

Construct validity provides theoretical and empirical evidence for the claim that a test measures what it purports to measure. A construct is the abstraction of the trait or ability that is being measured. For the Moby.Read test, the construct being measured is Oral Reading Fluency in English, or the ability to read English text aloud for meaning in a fluent manner. As shown in section 2, standard definitions of Oral Reading Fluency include four components: Comprehension, Accuracy, Accurate Reading Rate, and Expression. The Moby.Read test measures all four components, thus encompassing the full definition of the test construct.

To provide additional evidence for the validity of the Moby.Read test construct, AMI conducted two concurrent validation studies that provided empirical data showing that Moby.Read test scores closely correlate with scores of two other instrument purporting to measure the same or a similar construct: The Teachers College Running Records assessment and the Dynamic Indicators of Basic Early Literacy skills (DIBELS) NEXT assessment (see sections 7.3 and 7.4).

7.3 Concurrent Study of Reading Level

AMI conducted a concurrent study that compared Moby.Read Levels with independently assigned levels of a Teachers College Running Records assessment (TCRWP, 2018).

Participants. A total of 20 students in a Northern California school district took part in the study. Demographic information was not reported to protect student privacy.

Test Forms. A set of 21 forms was created for the study. Passages were assigned a Moby.Read Level based on the text leveling approach described in section 5.3.3). Three passages at each level were assembled into a form and named accordingly.

Procedure. Two facilitators administered the Moby.Read tests while students waited for the school's reading specialist to administer the Teachers College Running Records assessment. For the Moby.Read session, a facilitator presented the student with three word lists on paper and had the student read the lists out loud. Then the facilitator played an instructional video explaining the tasks and showing a student reading out loud with confidence. Based on *a priori* recommendations of reading level, the facilitator selected a form at the student's grade level and administered the test to the student. After the student completed the form, the facilitator looked up the student's Moby.Read Accurate Reading Rate score (which was the score automatically available at the time of the study) and selected another form of the appropriate difficulty level. For students performing well below the target range, the facilitator either ended the session or selected a follow-up form that was two levels lower than the form just read.

In all cases, the facilitator ended the session if the student showed fatigue. Across students, the number of forms presented ranged from 1 to 7, with an average of 3 forms per student.

Analysis. The data was used to create a model for empirical passage difficulty and final Moby.Read Levels (see section 6 Scoring). As a part of this model, raw ability scores for readers were generated and mapped onto Moby.Read Levels. The Moby.Read Levels were then correlated with independently derived Teachers College levels provided by the district's reading specialist.

Results. Moby.Read Accurate Reading Rate scores from 17 participants were compared to Teachers College levels. The correlation between scores of the two instruments was 0.93. Teachers College levels were not reported for three students because the students left before completing the test. Figure 4 shows a scatterplot of the scores for each of the participants.

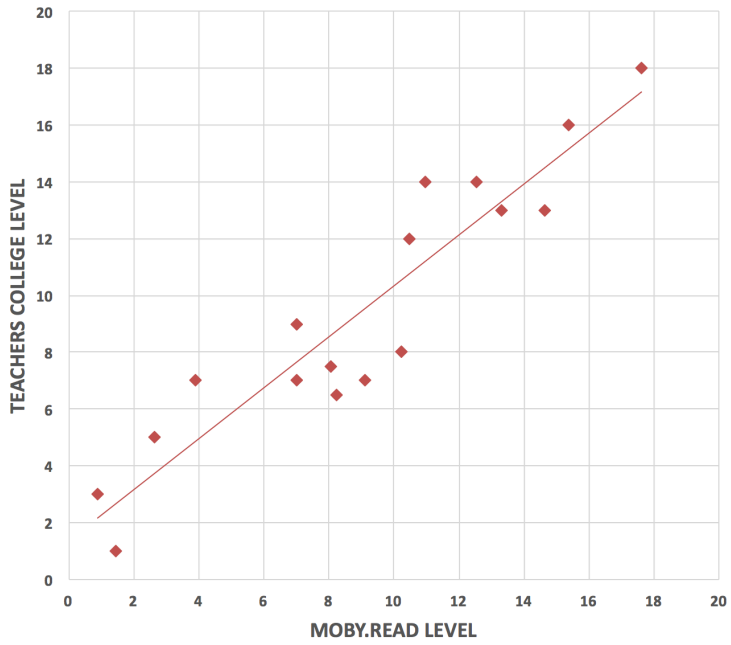

Figure 4: Scatterplot of Teachers College levels compared with Moby.Read Levels (n=17, r=0.93)

The high correlation between Moby.Read Level and TCRWP levels and the fact that the TCRWP's running record assessment measures the same construct of oral reading fluency provides evidence of construct validity for the Moby.Read assessment.

7.4 Concurrent Study of Accurate Reading Rate

A second study compared Moby.Read Accurate Reading Rate scores with scores of the Dynamic Indicators of Basic Early Literacy skills (DIBELS) Next assessment (Dynamic Measurement Group, 2018), a widely used oral reading fluency assessment.

Participants. Twenty students from an elementary school in California participated. Nine students were female, 11 were male. Seven were in 2nd grade, six were in 3rd grade and seven were in 4th grade.

Procedure. Students were given both a Moby.Read assessment and a DIBELS Next assessment. For half the participants, Moby.Read was administered first, and for the other half DIBELS was administered first. The administrator was an assessment professional with experience assessing reading within the DIBELS framework.

For the Moby.Read assessment, each student was fitted with a GearHead microphone headset. Students were administered a form appropriate for their grade. Moby.Read was delivered on an iPad Mini. The Moby.Read assessment was self-administered and automatically scored.

For the DIBELS assessment, students were administered a fall benchmark form consisting of three grade-leveled passages of about 250 words in length. The administrator followed the administration and scoring procedures described in the test's official documentation (Good, Kaminski, Cummings, et al., 2011). Students were given a passage and asked to read it out loud using the instruction prompts from the DIBELS assessment manual. The administrator started a timer when the student started reading. While the student read the passage, the administrator marked reading errors on a scoring sheet. After one minute, the timer beeped and the student's place in the passage was marked on the scoring sheet (none of the students

finished the passage). Then the student was asked to tell the administrator about the story. Responses were timed for one minute and marked on a comprehension scoring sheet, which simply tracked the number of words spoken in the student's response.

At the end of the session, the administrator pointed out that the student had done both a test on the iPad ("Moby.Read") and on paper ("teacher administered") and asked which the students preferred, and why. Consistent with standard practice, after the session, the administrator used the scoring sheets to calculate errors and Words Correct per Minute (WCPM).

Results. Moby.Read Accurate Reading Rate scores from 20 participants were compared to scores from the ORF task of DIBELS NEXT. The correlation between the two scores was 0.88 (Figure 5). Published studies investigating DIBELS report a test-retest reliability of 0.82 and an inter-rater reliability of 0.85 (Goffreda & DiPerna, 2010). The reliability of an instrument limits the strength of the correlation between that instrument and others measuring the same construct. So, the correlation with Moby.Read is at the ceiling of what would be expected given the reliability of DIBELS.

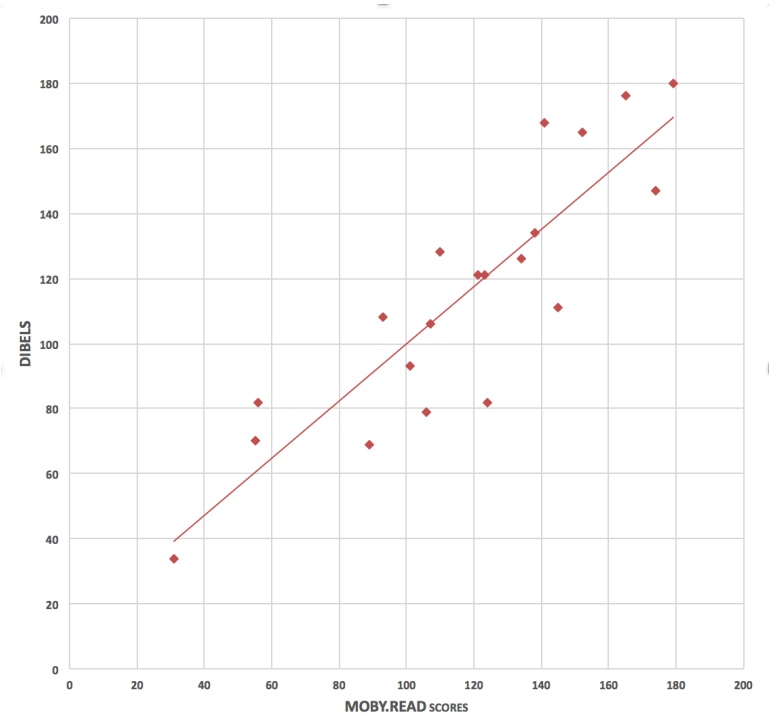

Figure 5: Scatterplot of DIBELS scores compared with Moby.Read Accurate Reading Rate (n=20, r=0.88)

7.5 Validation of Machine Scores

Moby.Read machine scoring algorithms were validated in a multi-site study that compared machine-generated scores with human ratings. Several analyses were performed on the resulting data to validate Moby.Read score components.

Participants. Participants in the study were 99 students from four different elementary schools. Participating schools included one public and one parochial school in New Jersey, and two public elementary schools in California. The female to male ratio was 47:52. Ages ranged from 7 to 10 with an average age of 8. Students were enrolled in 2nd Grade (29%), 3rd Grade (40%) and 4th Grade (31%). Regarding ethnic background (using classifications set forth by the US

Census), 51% of the students were European American, 19% were African American, 4% were Asian American, and 25% were identified as Hispanic or Latino.

Procedure. Two facilitators assisted with test administration: one in New Jersey and one in California. Both facilitators were assessment professionals. The experimental sessions with student-participants were conducted during the normal course of a school day at the participant's elementary school.

In preparation for the experimental sessions, the facilitators set up two or three chairs in a quiet area of the room or just outside the classroom. The Moby.Read assessment was delivered on an iPad Mini. Before the assessment, each student was fitted with a set of GearHead headphones with an inline microphone (the microphone was incorporated into the headset wire). Facilitators were present to help with technical problems, but they did not help students take the Moby.Read assessment. If a student asked a question during the assessment, the facilitator encouraged the student to keep going.

For all graded items, responses were machine scores and hand-transcribed. Readings were human-rated for Expression. Retelling responses and short-answer questions were human-rated for Comprehension. Following the test administration, students were presented a brief usability survey.

Analysis. Five participants were screened out because their test responses were silent or completely unintelligible. Data from 94 participants was used for further analysis.

7.5.1 Accurate Reading Rate

To validate Moby.Read Accurate Reading Rate scores, two analyses were performed. The first compared machine-generated Moby.Read scores with scores generated from human transcripts. At the passage level, the correlation between human and machine scores was 0.96.

In a second analysis, each recording of a passage reading was analyzed independently by two expert raters. Each rater listened to recorded student readings, marked errors, and measured the length of time the student read. Raters computed the WCPM for the passage by dividing the number of words read correctly over the duration of time of the reading (in seconds) times 60 (to place the units in minutes). Inter-rater reliability in this task was 0.99. Averages of the human-computed WCPM values were computed and the median score for each 3-passage session was derived for each participant. The median values of human scores were correlated with the median values of machine-generated WCPM scores. The resulting session-level correlation coefficient was 0.97.

Results confirm that the median Moby.Read Accurate Reading Rate scores are comparable to median values of scores produced by human raters. Figure 6 presents a scatterplot of machine versus human scores at the session level. The correlation of median WCPM scores between the two scoring methods suggests that scoring based on automatic speech processing and alignment with text has a high degree of accuracy.

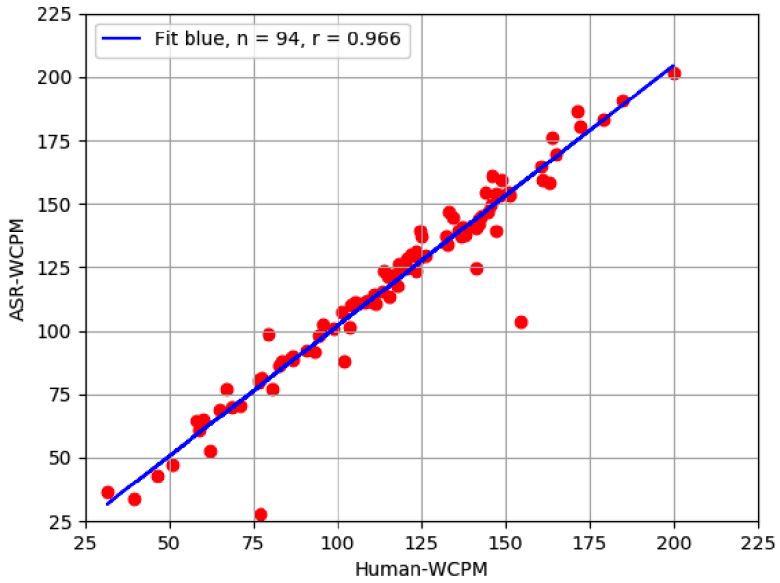

Figure 6: Scatterplot of Moby.Read Accurate Reading Rate compared with human ratings

7.5.2 Comprehension

To validate Moby.Read Comprehension scores, machine-generated Comprehension scores of passage retelling responses were compared with human ratings of the same material. Average machine scores were generated for each participant and were correlated with average human ratings. The human-machine correlation coefficient was 0.92. This correlation was better than that of the average human-human inter-rater correlation of 0.88. A scatterplot of the scores is shown in Figure 7.

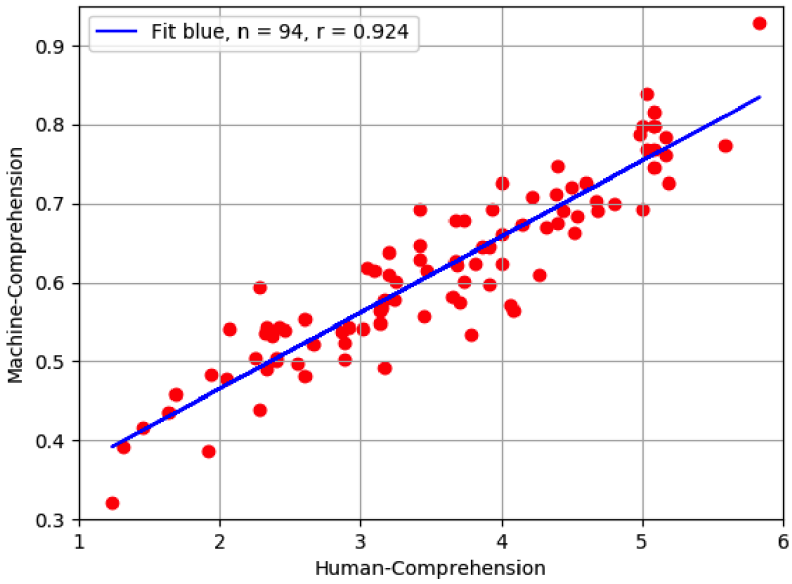

Figure 7: Scatterplot of Moby.Read Comprehension scores compared with human ratings

7.5.3 Expression Scores

To verify Expression scores, Moby.Read scores were compared to an average of three human ratings of Expression. For the three pairs of human raters, the average inter-rater correlation at the response level was 0.74. The correlation coefficient of machine-generated Expression scores and average human ratings of Expression was 0.88 (0.94 at the session level; see Figure 8), a statistically significant improvement. These correlations indicate that machine scores were more reliable in measuring Expression consistently than human raters.

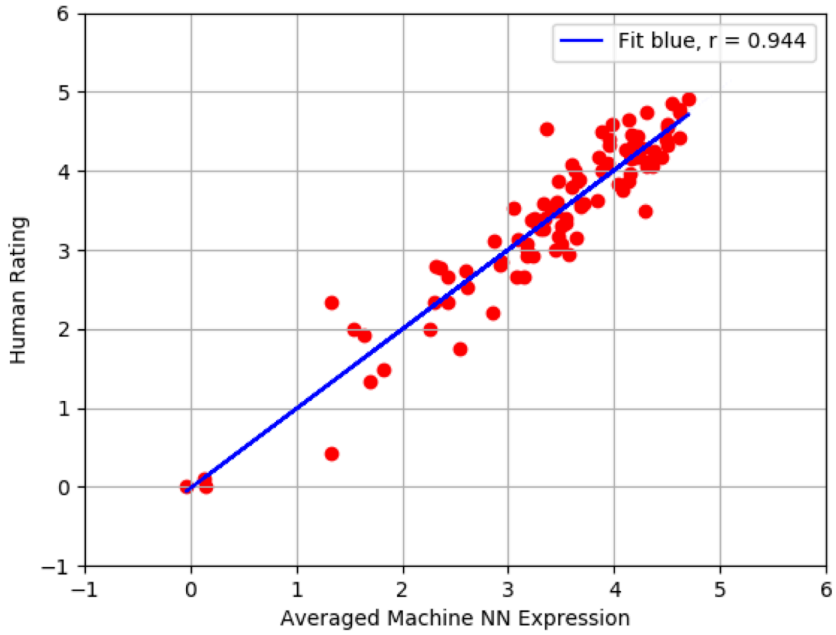

Figure 8: Scatterplot of Moby.Read Expression scores compared with human ratings

7.6 Usability

Usability data was collected at the end of the Moby.Read assessment for the same participants whose data were used to validate machine scores. Students were presented with a choice of image (shown in Figure 9) and asked to tap the screen image that best represented their experience with the app.

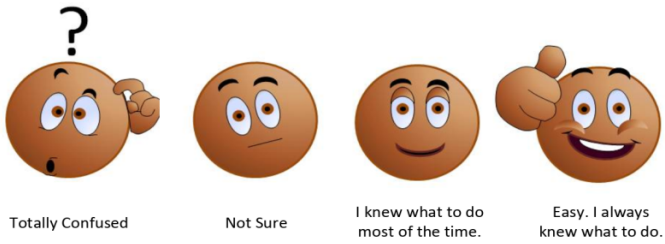

Figure 9: Four-point student usability rubric

Two students did not respond to the request for evaluating their experience. Among the remaining 97 who responded, 49% (n=48) selected *Easy. I always knew what to do.*, 43% (n=42) selected *I knew what to do most of the time*., and 7% (n=7) selected *Not sure*. No

student selected *Totally Confused."* Results show that 93% of students knew what to do most or all of the time, suggesting self-administration is viable for the majority of students.

A more detailed qualitative user survey was given to students at the end of the *Concurrent Study of Accuracy Reading Rate* (section 7.4). When asked which experience they preferred, 18 out of 20 students (90%) said they preferred "Moby.Read", and two students said they preferred "both"; not a single student preferred the teacher administration. Useful qualitative information was provided by the students when asked why they preferred the Moby.Read administration. The feedback is summarized in Table 15.

Table 15: Moby.Read student feedback

Theme	Student feedback
Technology	"I like screens" "There's this Siri thing"
Questions	"It asks questions, so you're reading for purpose" "You get to answer questions"
Re-read option	"I can read the stories again" "You could read it again"
Administration	"The iPad tells you what to do" "It tells you about the story before you read it"
Privacy	"It's more private" (i.e., no teachers are watching and judging)
User interface	"You have more time, so you can finish the story" "More pictures and stuff"

The qualitative user data collected suggests that a fully digital and automated oral reading fluency assessment that allows students to self-administer the test produces an engaging user experience for students.

8. Conclusion

Several forms of evidence support the validity and utility of the Moby.Read instrument.

Moby.Read passages sample a broad range of texts appropriate for students in grades 1 through 5. Validation studies provide evidence that the Moby.Read assessment accurately measures core components of oral reading fluency—Accuracy, Accurate Reading Rate, and Expression—while providing a measure of reading Comprehension that helps teachers accurately identify a student's reading Level. Results indicate that Moby.Read scores are comparable to, and correlate highly with, human ratings. Moby.Read scores for Comprehension and Expression have shown to correlate higher with average human ratings than scores from individual human raters with each other. Empirical data also provides evidence of construct validity with a high correlation between Moby.Read scores and fluency scores from other standardized assessments of oral reading. In sum, evidence supports the use of Moby.Read scores as a valid measure of oral reading fluency.

Students found the voice-interactive, automated assessment engaging and easy to use. The Moby.Read test structure and instruction format have been developed iteratively through multiple user studies and feedback. Pilot studies demonstrate that the Moby.Read user experience was sufficiently self-explanatory for 93% of students to self-administer the test successfully without teacher intervention. Further, 90% of students indicated they enjoyed using the Moby.Read application more than traditional oral reading fluency assessments administered by a teacher.

9. Analytic Measures Inc

Founded in 2014 and based in Palo Alto, California, Analytic Measures Incorporated (AMI) is an employee-owned corporation founded on the principle that strong data science and technology applied to educational tools can enhance and accelerate learning. AMI designs and builds advanced artificial intelligence (AI) applications and machine learning technologies for education products. AMI applies machine learning, psycholinguistics, and psychometrics to products and scoring systems. AMI uses automated speech and text evaluation technologies to measure a range of skills, including reading, communication, academic content, and social/emotional skills. For more information, visit www.analyticmeasures.com.

10. References

Brown, L. T., Mohr, K. A., Wilcox, B. R., & Barrett, T. S. (2017). The effects of dyad reading and text difficulty on third-graders' reading achievement. *The Journal of Educational Research*, 1-13.

Cheng, J. (2018). Real-time scoring of an oral reading assessment on mobile devices. INTERSPEECH 2018, September 2-6, Hyderabad, India. Retrieved from https://www.isca-speech.org/archive/Interspeech_2018/pdfs/0034.pdf

Council of Chief State School Officers (CCSSO) & National Governors Association (NGA). (2012). Supplemental information for Appendix A of the Common Core State Standards for English language arts and literacy: New research on text complexity. Washington, DC: CCSSO & NGA.

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

Dynamic Measurement Group. (2018). DIBELS Next. Retrieved from https://dibels.org/index.html

Educational Testing Service (2015). *Guidelines for Fair Tests and Communications*. Princeton, NJ: Educational Testing Service.

Kincaid, J. P. Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel. Research Branch Report 8-75. Chief of Naval Technical Training: Naval Air Station Memphis.

Goffreda, C.T. & DiPerna, J.C. (2010). An empirical review of psychometric evidence for the Dynamic Indicators of Basic Early Literacy Skills. *School Psychology Review, 39*(3), 463-483.

Good, R. H., Kaminski, R. A., Cummings, K., Dufour-Martel, C., Peterson, K., Powell-Smith, K., & Wallin, J. (2011). *DIBELS next assessment manual.* Eugene, OR: Dynamic Measurement Group.

Hasbrouck, J. & Tindal, G. A. (2006), Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers. *The Reading Teacher,* 59: 636–644.

Herman, P. A. (1985). The effect of repeated readings on reading rate, speech pauses, and word recognition accuracy. *Reading Research Quarterly*, 553-565.

Johnson, M.S. & Kress, R.A. (1965). Informal reading inventories. *IRA Service Bulletin.* Newark: Del.: International Reading Association.

Jurafsky, D. & Martin, J. H. (2009). *Speech and language processing*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Miller, J. & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly, 43* (4), 336-354.

Morgan, A., Wilcox, B. R., & Eldredge, J. L. (2000). Effect of difficulty levels on second-grade delayed readers using dyad reading. *The Journal of Educational Research*, *94*(2), 113-119.

Morrow, L. M. (1985). Retelling stories: A strategy for improving young children's comprehension, concept of story structure, and oral language complexity. *The Elementary School Journal*, *85*(5), 647-661.

National Assessment Governing Board, U.S Department of Education. (2017). *Reading framework for the 2017 National Assessment of Educational Progress (NAEP).* Washington, D. C.: NAGB.

National Governors Association (NGA), Center for Best Practices & Council of Chief State School Officers (CCSSO). (2010). *Common core state standards (English language arts, reading foundational skills, fluency)*. Washington, D.C.: National Governors Association Center for Best Practices & Council of Chief State School Officers.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K. (2011). The KALDI speech recognition toolkit. In proceedings of *IEEE Workshop on Automatic Speech Recognition and Understanding.*

Rasinski, T. V., Padak, N. D., McKeon, C. A., Wilfong, L. G., Friedauer, J. A., & Heim, P. (2005). Is reading fluency a key for successful high school reading? *Journal of Adolescent & Adult Literacy, 49*(1), 22-27.

Teachers College Reading and Writing Project. (2018). Running records assessment. Retrieved from https://readingandwritingproject.org/resources/assessments/running-records

Texas Education Agency. (2017). Texas Essential Knowledge and Skills for English Language Arts and Reading. Retrieved from https://tea.texas.gov/curriculum/teks/

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education*, *25*(4), 252-261.

Wilson, R. M., Gambrell, L. B., & Pfeiffer, W. R. (1985). The effects of retelling upon reading comprehension and recall of text information. *The Journal of Educational Research*, *78*(4), 216-220.

Zhang, X. Trmal, J., Povey, D. & Khudanpur, S. (2014). Improving deep neural network acoustic models using generalized maxout networks. Proceedings of *ICASSP,* 215-219.

Appendix: Moby.Read Text Types

Table 16: Text types per passage and grade level

Grade Level	Passage	Type
1st Grade Fall	First Passage (#1401)	narrative
	Second Passage (#1603)	informational
	Third Passage (#1903)	narrative
1st Grade Winter	First Passage (#2001)	narrative
	Second Passage (#1502)	informational
	Third Passage (#1703)	narrative
1st Grade Spring	First Passage (#1505)	narrative
	Second Passage (#1807)	informational
	Third Passage (#1804)	narrative
2nd Grade Fall	First Passage (#2401)	narrative
	Second Passage (#350)	informational
	Third Passage (#2203)	narrative
2nd Grade Winter	First Passage (#2302)	narrative
	Second Passage (#2102)	informational
	Third Passage (#310)	narrative
2nd Grade Spring	First Passage (#330)	narrative
	Second Passage (#2201)	informational
	Third Passage (#2202)	narrative
3rd Grade Fall	First Passage (#2701)	narrative
	Second Passage (#420)	informational
	Third Passage (#2403)	narrative
3rd Grade Winter	First Passage (#410)	narrative
	Second Passage (#2504)	informational
	Third Passage (#2101)	narrative
3rd Grade Spring	First Passage (#450)	narrative
	Second Passage (#2503)	informational
	Third Passage (#320)	narrative
4th Grade Fall	First Passage (#520)	narrative
	Second Passage (#2402)	informational
	Third Passage (#2904)	narrative
4th Grade Winter	First Passage (#2803)	informational
	Second Passage (#2902)	informational
	Third Passage (#3106)	narrative
4th Grade Spring	First Passage (#2801)	informational
	Second Passage (#3105)	narrative
	Third Passage (#2601)	informational
5th Grade Fall	First Passage (#3003)	informational
	Second Passage (#2901)	narrative
	Third Passage (#3002)	informational
5th Grade Winter	First Passage (#2702)	informational
	Second Passage (#3401)	narrative
	Third Passage (#3202)	informational
5th Grade Spring	First Passage (#2802)	informational
	Second Passage (#3201)	narrative
	Third Passage (#3101)	informational