#### PART OF A PHASE I R&D PROPOSAL TO THE U.S. Dept of Education

[written for an education reader with little knowledge of CS]

#### Intro

Reading is an essential skill for learning and for civic engagement. Nevertheless, in the 2017 representative sample of the 3.8 million U.S. 4th-graders, NAEP found that 32% of the students are “below basic” in reading (NCES, 2017). Further, there is a significant cost to the student and to the nation when reading difficulties are not diagnosed until 2nd or 3rd grade, which delays the formulation and application of interventions. Students who lag behind their cohort in reading in 3rd grade often show significant long-term deficits in all areas of education (NRC, 1998). For these reasons, many hours of teacher, parent, and student time are invested in practicing reading skills, taking mandated and elective assessments of early reading, and implementing intervention strategies to accelerate skill acquisition in reading. In grades K-3, these assessments are often Oral Reading Fluency (ORF) measures for fall-winter-spring benchmarking and for struggling readers the assessments are more frequent short tests to monitor progress in response to a reading intervention.

### a. Problem

The problem is that the rich information produced by students reading out loud during early elementary assessments is usually not recorded and not analyzed to guide instruction. An ORF test score may indicate a student is reading below grade level expectation, but the assessment usually provides no formative guidance for intervention and remediation. Skilled teachers and reading specialists who observe and listen to a young student reading aloud can infer a profile of likely reading challenges to guide remediation, but they either have to perform a reading diagnosis in real time or record an ORF session and listen to it repeatedly to extract and code the evidence of specific areas of reading difficulty. Even then, consistent application of diagnostic criteria across students and across administrations is not easily scalable.

The problem is exacerbated for those teachers who do not have the training or skill or confidence to infer a reading profile from a read-aloud performance, no matter how many times they might hear a student read from authentic materials. For these teachers, a path forward may be to assign yet more tests to diagnose reading problems or to follow a general prescribed curriculum for this level reader. Assigning more tests puts a burden on students and teachers (and/or district specialists), while continuing with the general curriculum may leave the student’s specific difficulties unaddressed.

Analytic Measures Inc. (**AMI**) proposes to implement and evaluate a technical innovation that can automatically derive a profile of reading strengths and difficulties from recordings of a student’s performance in reading several short passages aloud. This technology promises to eliminate or greatly reduce the diagnostic testing burden and help every teacher focus on those reading difficulties that students actually encounter when reading grade-level prose.

### b. Moby.Read Profiler: its implementation and intended effect on outcomes

AMI’s **Moby.Read**(TM) system is an Oral Reading Fluency (ORF) instrument designed, built and validated with IES SBIR funding (2016-19). Moby.Read was introduced commercially in January 2019. See [www.analyticmeasures.com/moby-read](http://www.analyticmeasures.com/moby-read) for a product description and access to a demo of the current Moby.Read assessment, or see <https://youtu.be/_V6_7agY5tc> .

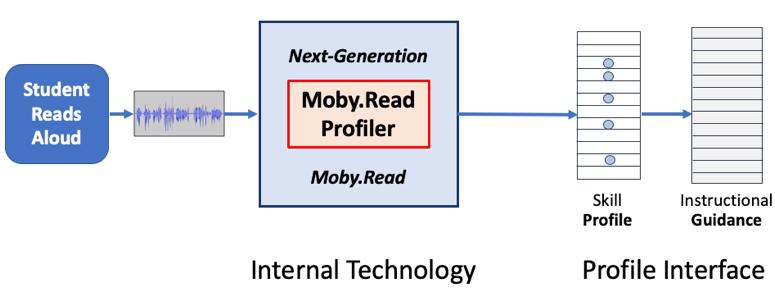
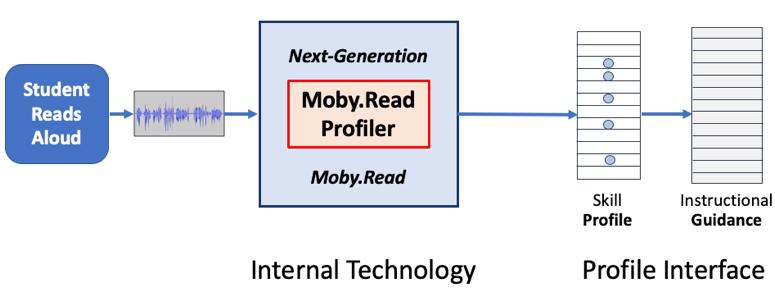
The usual dimensions of oral reading fluency are rate, accuracy, and expression, and Moby.Read already automatically reports these measures with high accuracy from a student’s reading of three short passages. Moby.Read augments these fluency measures with comprehension measures based on spoken retellings and constructed answers to short questions. AMI has developed spoken language processing (SLP) modules and scoring algorithms that have achieved excellent accuracy in measurement of oral reading rate, accuracy and expression. The SLP modules have been combined with AMI’s natural language processing (NLP) algorithms to measure reading comprehension from spontaneous spoken retellings and spoken answers to comprehension questions. The new step is to recombine these technologies and apply them to extract the underlying skill profile that determines a student’s oral reading fluency and comprehension performance.

The proposed project will develop a Moby.Read Profiler (**MRP**) technology that can power a fuller diagnostic extension to AMI's Moby.Read oral reading fluency assessment service. The Profiler function will still operate on oral readings of passage text, but it will provide a major addition to the reporting output that will enhance its usefulness for teachers.

At the end of a Phase II of this project, AMI should have a complete, reading-skill profiler integrated in the Moby.Read system, which will indicate suitable reading interventions to accelerate skill acquisition for successful reading. The accuracy of the profiles will have been validated, the student and teacher interfaces will have been refined in several iterations, and the Moby.Read Profiler’s usefulness to classroom teachers will have been demonstrated with appropriate studies of student outcomes.

As shown in Figure 1, the two goals of the Phase I project focus are:

1. Internal technology: demonstrate that AMI can automatically analyze oral passage readings to extract accurate scores on several dimensions of early reading skill, the measurement of which currently require specific diagnostic testing or a highly skilled human analysis; and

******2. Profile interface: evaluate the design of a teacher interface to support effective use of the skill profile information by ordinary teachers, try it out in classrooms, determine if teachers are comfortable with the profile presentation format and if they think the information will be useful for setting individual instructional plans, and collect data to specify the next iteration of the profile interface design.

The MRP’s intended users are early readers and struggling readers of any age. Students will use Moby.Read to capture samples of their oral reading performances and elementary school teachers will use MRP output to understand each student’s reading profile and to work with that student on suitable instructional materials. An authentic, 12-minute read-aloud and retelling of four short passages captures much of the information needed to identify areas of reading difficulty for that student.

**Figure 1**. Phase I research focus: internal technology and profile interface design

The MRP would function as shown in Figure 1. A person reads aloud from a text and the audio signal is captured and automatically analyzed to produce a matrix of performance data (see Figure 2 below). The text is also analyzed automatically to produce a matrix of text-analytic data. These two matrices are processed by a Performance Profile Model (PPM) that produces a reading skill profile. The PPM (described in section 2.a below) is the core technical result of this development project. From the reading skill profile, MRP will suggest generic best-practice interventions and published instructional resources.

Like the current Moby.Read system, the Profiler will be platform agnostic. As of 2019, Moby.Read runs in any recent Chrome browser, under Windows, MacOS, or Android. It also runs as a native app in iOS devices (Apple iPads and iPhones).

Moby.Read Profiler will be designed for all students in Grades K-5 and will be self-administered in school and classroom settings as part of the Moby.Read Service. Given its diagnostic value, the MRP will provide value to strong and struggling readers alike when passage reading performances directly enable reporting of key reading strengths, challenges, and potential interventions that support continued improvement.

After the students use Moby.Read to generate material for the profiler, their teachers will be the primary users of MRP, when they interact with Moby.Read’s new Profile Interface. As noted, the MRP will provide a reading profile and specific skills to focus on for each student as well as interventions aligned to instructional materials and best practices.

MRP is designed to be an extension of Moby.Read. Moby.Read is currently designed to make the administration of ORF benchmark assessments more efficient and standardized for teachers. The MRP will build on this design and afford teachers more targeted reading performance data without the time or financial burden of additional diagnostic assessments.

… …

### b. R&D objectives

The overall objective of the Profiler project is to demonstrate the technical feasibility of the Profiler concept through design and development of a prototype system, and then perform a preliminary evaluation of the Profiler prototype through a study of its use and utility in a classroom setting. The specific R/R&D objectives of the Phase I project are:

**1.** Design Skill Profile:

a. Identify and validate the key events that reading specialists hear in oral readings and which skills reading experts can consistently infer from events in oral readings.

b. Design a Profile Report for presentation of skill-profiles to classroom teachers.

c. Elicit input from elementary teachers on organizing the content of the Profile Report

**2.** Develop a Performance Profile Model that automatically predicts a skill profile from a student’s oral reading of known texts.

**3.** Integrate the Profiler into the Moby.Read system and implement the Profile Report within the Moby.Read teacher interface.

**4.** Pilot Moby.Read Profiler in authentic classroom settings; study use and utility for teachers.

The prototype Profiler will be built into the existing Moby.Read system and fielded as part of Moby.Read, along with the other main component, the Profile Interface. Both are needed to study the usability of the Moby.Read Profiler.

Going one level below, the critical sub components needed for development of the Profiler itself are shown in Figure 2. These subcomponents are:

1. Grade-appropriate texts for students to read. (available)

2. Audio recordings of student performances reading aloud. (available)

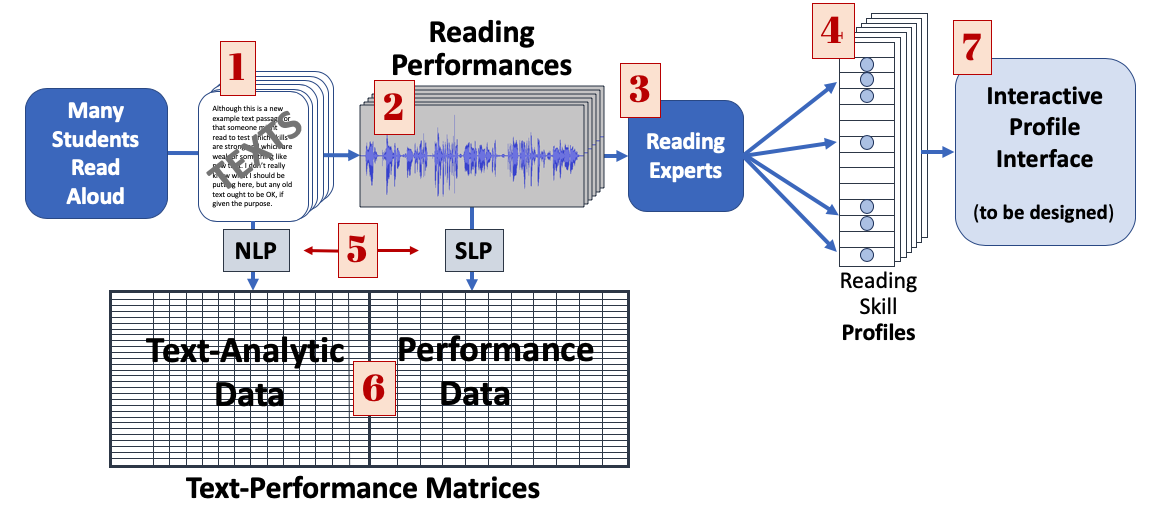
3. Reading-specialist evaluations of the reading performances. (available)

4. Reading skill profiles distilled from the reading-specialist evaluations & reviewed by experts.

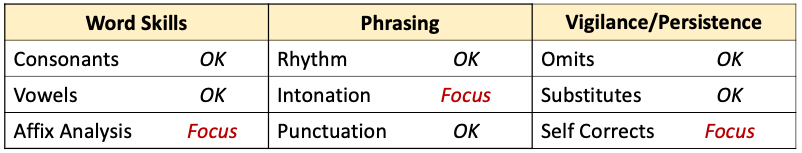
5. AI resources: NLP and SLP modules for analysis of texts and oral readings. (available)

6. Text-Analytic data matrices and oral reading Performance matrices. (data partially available)

7. Interactive Profile Interface designed for use by classroom teachers.

Among the seven subcomponents needed to build a Profiler for Phase I user studies, AMI already has four of the subcomponents {1, 2, 3, and 5} as well as a portion of the text and performance data (139 readings of 27 passages) that are needed to develop the Profiler. The first missing subcomponent is (4), the reading skill profile for each of the students whose data will be used as the target of the Profiler optimization. That is, the Profiler will be developed and optimized to produce student reading skill profiles that are as close as possible to the Profiles that have been derived from the same oral reading audio by reading specialists. Figure 3 shows an example of the information that may be in the Phase I profiles.

**Figure 2.** Seven subcomponents needed to build Phase I research prototype within Moby.Read.

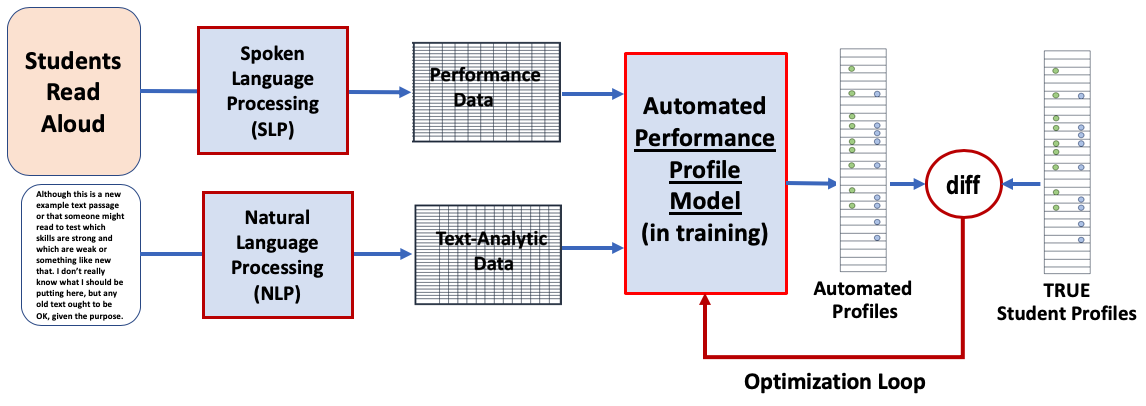
Subcomponent 4 depends on the two data matrices. The Performance Data matrix is produced by AMI’s SLP module, which currently produces an internal transcript of the words read aloud correctly and incorrectly, as well as the time the student takes to read each word and the slope of reading rate. AMI’s scoring system also counts the misreads (including substitutions and deletions), and it can be extended to track and categorize false starts. For several of the texts in the Moby.Read system, AMI’s NLP module has already produced the Text-Analysis Data matrices. Matrices will be produced for the remainder of the passages to be used in the Phase I study.

**Figure 3.** An example of possible Phase I profile information.

The core R&D task in Phase I is optimization of the automated skill Profiler. The Profiler processes two data matrices for each of four passage readings from a given student and produces a skill profile. The Text-Analysis Data provides word-by-word information on the passage structure and the task-relevant features of each word, including the word’s part of speech, its probability given preceding and following context, its length in letters, its expected mastery grade level, and two estimates of the word’s expected duration in context.

The Performance Data provides the observed time it took to produce each word or miscue, a count of false starts, and the “slope” of the reading rate. All this is in addition to the current output of Moby.Read: overall rate, percent word accuracy, expression and comprehension.

The information in these two data matrices, along with the text level and the student’s grade level is combined in the Profiler to produce a profile. AMI’s deep learning procedures (that produce a deep neural net) will process an existing data set that includes 32 passages, each read by 20 students (640 recorded reading performances). In parallel, a group of 2-3 reading specialists will write their observations and instructional suggestions for the same 640 recorded readings, producing, for each reading, double annotations on the performance. From the resulting 1280 annotations and suggestions, AMI (with WestEd) will distill a set of plausible oral-reading descriptors and related instructional suggestions that should be predictable from the text structure and a student’s reading performances on four passages. These three data sets (performance data, text-analytic data, and “true” reading-specialist profiles) are the input to the process that optimizes the Profiler’s output to match the corresponding specialist profiles.

The iterative procedure to optimize the combination of the two matrices is shown in Figure 4. The loop that trains the Performance Profile Model uses information in the performance data and text-analytic matrices to produce automated profiles that predict the true student profiles as they were determined by combining specialist judgments. The optimization will be run on about 80% of the student data and then its accuracy will be estimated on its match to the remaining 20% of the data that will be held out as a test set.

**Figure 4**. Optimization of the Performance Profile Model by iterative machine learning. A deep neural net is trained to match a sample of reading profiles.

… …

### c. Potential problems

AMI does not anticipate major problems with the prototype development or the usability and feasibility study during Phase I. However, some of AMI’s most effective machine learning techniques work best with very large data sets, whereas the Phase I data sets (comprising 640 matrix pairs and 1280 annotation texts) are relatively sparse to estimate 6-9 variables. If the machine learning process does not converge on a reasonable set of variables, the data set is small enough to apply semi-automatic methods to reduce the variability in the free-form annotations, or to impose a dimensional structure and a fixed rubric set on the specialist raters.

## References

J. Balogh, J. Bernstein, J. Cheng, A. VanMoere, B. Townshend, & M. Suzuki (2012) ”Validation of automated scoring of oral reading”, *Educational & Psychological Measurement*, 72(3), pp. 435-452.

J. Bernstein, J. Cheng, J. Balogh, & R. Downey (2019) Artificial intelligence for scoring oral reading fluency. A chapter in *Applications of artificial intelligence to assessment*, H. Jiao & R. Lissitz, Eds. Info Age.

J. Bernstein, A. VanMoere & J. Cheng (2010) “Validating automated speaking tests” *Language Testing* 27(3) 355-77.

J. Cheng, Y. D’Antilio, X. Chen, & J. Bernstein (2014) “Automatic assessment of the speech of young English learners”. Paper at the BEA workshop, NAACL-HLT 2014, June, Baltimore.