Research Design and Methods

The proposed CMARS is conceived as a web-based, computer administered assessment tool designed to be both student and teacher friendly. It will be student friendly in that each assessment session will feel to the student like he or she is playing a fast paced computer game called, "Right Stuff University." The student will be engaged in a serried of fast-past, ability tailored subtests requiring no more than 30 minutes to complete in total. After the completion of each subtest, student will be incentivized to continue to achieve peak performance by the sharing of performance data and goal setting. It will be teacher friendly because it will be computer administered, thus requiring little time commitment for the teacher; and yet, will provide immediate, easily interpretable information about student progress. Further, teacher will receive immediate class wide feedback to assist with grouping and instructional target decisions, as well as link to downloadable lesson plans specific to target skills.

Phase I Research Design and Methods

The purposed CMARS assessment will be comprised of five subtests representing the four domains of reading, previous identified. The domain of Word Analysis will be assessed through the spelling subtest. The domain of Fluency will be measured through the connected text fluency and silent reading fluency subtests. The domain of Vocabulary will be measured by the vocabulary subtest, and will include included both general and content area vocabulary. The domain of Comprehension will be measured by the comprehension subtest, and will include several types of comprehension abilities including: determining main idea, making inferences, making critical judgmentrdpk tehensioilitincludhefkeff will lnformshie isi.s.

aspects of fluency. The Maze task has been shown to be highly correlated to measures of both fluency and comprehension and has high reliability and concurrent validity (Brown-Chidsey, Davis, & Maya, 2003; Fuchs & Fuchs, 1991; Jenkins, Pious, & Jewell 1990; Shinn, Good, Knutson, Tilly, & Collins, 1992; Swain & Allinder, 1996). A similar task was part of the CMERS assessment. Our data confirms that our Maze task, delivered via computer correlates highly to measure of oral reading fluency, comprehension measures, as well as high stakes assessments (Kalinowski, 2009).

Procedure. To complete connected text fluency, the computer will tell students it is time to read a story and review the procedures. The first page then appears, and students perform the Maze task for two and one-half minutes, or until they complete the story. When students complete a page, they click on a button to turn the page and continue. The score obtained from this incorporates the number and accuracy of Maze items completed in the allocated time, as well as accounts for the number of words read between Mazes. This score, which our team formulated for CMERS, has been shown to better correlate to other measures of both DIBELS Oral Read Fluency and comprehension (Lyon & Kalinowski, 2008). An example Maze task authored in the CMARS space exploration theme appears in Figure 4.

Figure 4. Sample Maze task for CMARS.

Silent reading fluency subtest. A second aspect of fluency is a student's ability to silently read connected text that is matched to their decoding ability. We propose to measure silent reading fluency by asking students read passages written at a level they can decode while being timed by the computer. Students complete the silent reading of a passage, then press an icon to answer questions about the passage (see the comprehension subtest below). Students will read both narrative and expository passages. While the lower grades will see an equal combination of these two types of text, the upper grades will be reading more expository passages than narrative text. Passages will be composed of varying word counts of 250 to 500 words, with passages written at lower levels being shorter, and more advanced passages longer. There will be 220 total passages created of varying complexity and difficulty, ranging in readability of 2.0 through 12.9 on the Flesh-Kinkaid scale. To assist teachers with assessing the reading ability of their students, we will also Lexile each passage.

Theory and research. Students at about grade 4 transition from gaining more meaning from text read orally to gaining more meaning from text read silently (Prior & Welling, 2001). Not surprisingly, the correlations between traditional Oral Reading Fluency measures and other aspects of reading also become weaker at this time (Brown-Chidsey et al., 2003). Since the ability to read text silently takes on greater importance, and because students in the grades beyond grade 4 are expected to read most of their text silent, the importance of assessing students' ability to read text silently with fluency cannot be overstated. Even so, in our review of the literature, we found that little attention has been paid to this important aspect of reading. Thus, the inclusion of silent reading fluency is, in many ways, experimental. Through the proposed work, we will be able to determine: (a) if silent reading fluency is amenable to measurement in the way that propose, and (b) how well it correlated to the other more established measures of fluency and comprehension. An important aspect of this proposed measure is that we are placing students into text for which they have the ability to actually read the words comprised in the text. Thus, we will be able to ascertain students' silent fluency on text for which they possess the ability to decode.

Procedures. For this subtest, the computer will announce that it is time to read a passage and answer questions (see Figure 5). Students will be told that the computer is timing them as they read the passage, but that they need to read the passage carefully enough to understand the passage without returning to the text. Timing will begin when the passage appears on the page and will end when the student turns the page to begin answering comprehension questions regarding the content of the passage.

Figure 5. Sample silent reading fluency / comprehension passage for CMARS.

Vocabulary subtest. Students will demonstrate their knowledge of word meanings through synonyms or definitions, as well as the ability to infer meaning through context. Four types of questions will be used: (a) select the word that best matches the following definition, (b) select the word that is most similar in meaning to the following word, (c) select the word that best describes the following picture, and (d) select the word that is most similar in meaning to the underlined word. Distracters choices for each word will include words with a similar spelling or pronunciation, antonyms, words an unrelated meaning.

Theory and research. In order to assess students' knowledge of word meaning, we will use decontextualized type of items (synonyms, picture, and definition). However, we also know that students acquire vocabulary best when it is used in a meaningful context. Thus, we also include contextual type of questions, in which students must infer the correct meaning of a word based on its use in a sentence. We have chosen passive recognition tasks for our assessment based on reports that the ability to establish the link between word form and word meaning is the most important component of word knowledge (Laufer et al., 2004; Read, 2007)

Procedures. Throughout the vocabulary assessment, there will be a mix of general vocabulary words and content vocabulary words. The narrator will read the stem for each item. Students can choose to hear the word choices by scrolling over each word on the screen. Students will choose among four possible answers by clicking their mouse on their selected answers. See Figure 6 for an example. The computer CAT program will match the difficulty of the items to the abilities of the students regardless of their age or grade level. Teachers will be able to access reports of their students' progress and needed areas of vocabulary instruction.

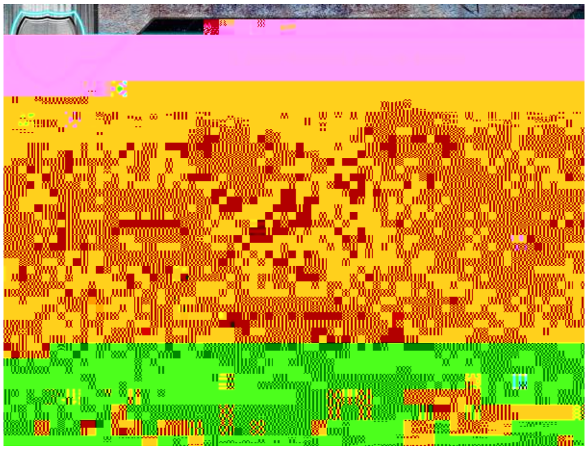

Figure 6. Example vocabulary item for CMARS.

Reading comprehension subtest

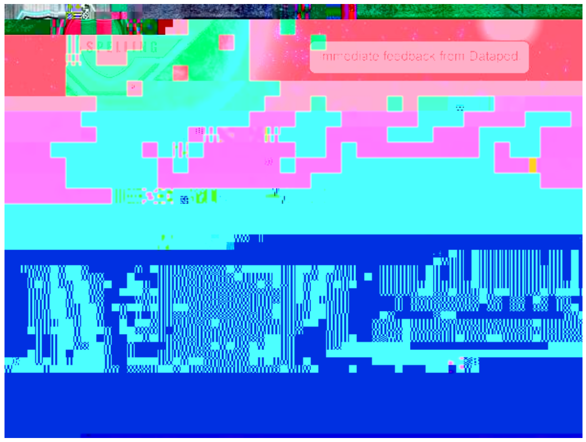

Figure 8. Sample student feedback for CMARS.

While the graph for each subtest is the only feedback information shown to children, both graphic and skills analysis information is provided to the teacher with an accounting of (a) the skills for which the child has already demonstrated mastery, (b) the skills on which the child is being assessed, (c) performance details on the skills being assessed. Scores reported will include: (a) an IRT-based ability index that represents an estimate of the child's absolute level of ability in a given domain. Because it is not restricted to age or grade levels the ability index can be used to show growth over time, (b) a relative class score representing the quartile range the child is in for her class (i.e. bottom 25%, 25th to 50th percentile, 50th to 75th percentile, and top 25% of the class), and (c) a normed-based percentile ranking comparing the ability scores to a large, nationally representative sample.

Class level data. Because this software is being designed for class wide implementation, the existing CMERS data management and reporting system will be utilized, allowing teachers to examine both individual child and classroom level data. Each individual child's data is recorded in a classroom file. This will allow the computer to aggregate data in the class and generate class level reports. Data for the class will be displayed in rank order form using the most recent data. This is presented in four columns: Mastered, Above the Mean, Below the Mean, and Not Yet Included. Under each column, the names of appropriate children are listed in rank order from highest to lowest. From this list, a teacher can automatically transfer to an individual child's file by clicking on the child's name. For children who are included in the subtest, the child's most recent score appears next to their name. Next to the score appears the child slope or trend score. Children whose slopes are near zero or negative are highlighted to alert the teacher to attend to their academic needs. Likewise, if a child has been designated as requiring "intensive care," the word help appears in red letters next to that child's name. For a child to be designated as requiring intensive care, he or she have must be scoring in the bottom quartile based on our norm reference sample on two consecutive assessments and have slopes are near 0 or negative.

Higher level data. Likewise, a data-gathering feature that will aggregate the data from several classrooms will be developed based on existing code from CMERS. We see this feature as most pertinent to district and building level personnel such as principals, reading specialists, school psychologists, grade level lead teachers, or language arts coordinators. Since CMARS, like CMERS, is web-based, aggregated reports can be generated for virtually any desired aggregation level (nation, state, district, building, or grade) and by any sub-population (gender, ethnicity, SES status, etc.).

establish item-level parameters. Subsequently, the items were programmed into a CAT framework for commercialization. In 2008-09, CMERS was delivered in a controlled study to over 400 kindergarten through grade 3 students, along with well-regarded measures of reading ability, to establish reliability and validity evidence (Kalinowski, 2009). Since then, CMERS has been well received by teachers, districts, and state agencies as a respected instrument for continuous progress monitoring of early reading skills. We will use the experiences gained from commercializing CMERS to successfully bring CMARS to the market.

IRT calibration study. An IRT calibration study will be used to determine the item parameter estimates for the pool of items used with CMARS. As with CMERS, a two-parameter logistic (2PL) model will be used to allow for both the item difficulty parameter, as well as the item discrimination parameter to vary by item. Equation 1 illustrates the 2PL model predicting the probability of a correct response to item *j*:

$$P(X_j \mid \theta, a_j, b_j) = \frac{1}{1 + e^{-a_j(\theta - b_j)}}, \quad (1)$$

where θ is the person location parameter (i.e., ability), and b_j and a_j are item *j*'s location parameters (i.e. difficulty and discrimination parameters, respectively; Lord, 1980). The 2PL model will be used for the item types having dichotomous responses, such as vocabulary items. For item types with polytomous responses (i.e., correct, partially correct, and incorrect), such as spelling and comprehension testlets, we propose to use Muraki's (1992) generalized partial credit model, also known as the two-parameter partial credit model (2PPC) model (Yen, 1993) as a natural extension to the 2PL model.

Research design. To determine item-level parameters as well as address the model assumptions, a nonequivalent multi-group IRT calibration study has been developed. Students will be recruited from local area schools in much the same way as with the previous CMERS study. However, given the increased number of items, across multiple grade levels (2-14), only a portion of items will be given to students at each grade level. See Table 1 for the proposed distribution of items to students.

Table 1
Nonequivalent, Multi-group Design for IRT Calibration

		Item difficulty (estimated grade level)						
		2-3	4	5	6	7	8	9-14
Students (actual grade level)	8					X	X	X
	7				X	X	X	
	6			X	X	X		
	5		X	X	X			
	4	X	X	X				

Note. An "X" represents a group of students taking items at a particular level of difficulty.

Our goal is to recruit approximately 400 students at each grade level (4-8), for a total of 2,000 students in the study. Given this design, each item will have between 400 and 1,200 responses, which is adequate for accurate parameter estimation in a 2PL IRT model (de Ayala, 2009). To estimate the item parameters, BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 2003) will be used for dichotomous item types, and MULTILOG (Thissen, Chen, & Bock, 2003) will be used for polytomous item types. Both IRT programs use marginal ml1Rcum likelihood estimation (MMLE) to ml1Rcize the person response vector across both the item difficulty and discrRcinability dimensions. For example, Equation 2 represents the probability of a response vector of dichotomous items, **X**, in an instrument of length L,

$$P(\mathbf{X} \mid \theta, \mathbf{J}) = \prod_{j=1}^{L} p_j^{x_j} (1 - p_j)^{1 - x_j} \ , \qquad (2)$$

where the probability of a set of responses is conditioned on the person's ability (θ) and the matrix of item parameters, **J** (i.e., the collection of *a*

ion mlikelihood bw-2

Reliability and validity study. A reliability and validity study will be used to determine the consistency and accuracy of CMARS data as compared to other widely used measures of reading ability.

Research design. Approximately 500 students from grades 4-8 will be recruited from multiple North Texas school districts to participate in the study. Students will be escorted to the school's computer lab and administered all assessments by trained graduate research in the data can found regarding areas of the reports that need to be addressed, changes will be proposed to engineering prior to commercialization.

In Phase II, the main goal will be to determine the psychometric properties of the pool of items developed and authored in Phase I, program them into a CAT framework, and collect reliability and validity evidence for the resulting CMARS instrument. Further, focus groups on select stakeholders will help istation determine if the existing reporting interface is satisfactory for making continuous progress monitoring decisions for their students. After completion of Phase II, CMARS will be ready for commercialization into established markets that have embraced its predecessor, CMERS.