Psychology in the Schools, Vol. 00(0), 2012
View this article online at wileyonlinelibrary.com/journal/pits
© 2012 Wiley Periodicals, Inc.
DOI: 10.1002/pits.21613

SILENT READING FLUENCY USING UNDERLINING: EVIDENCE FOR AN ALTERNATIVE METHOD OF ASSESSMENT

KATHERINE W. PRICE, ELIZABETH B. MEISINGER, AND MAX M. LOUWERSE

University of Memphis

SIDNEY K. D'MELLO

University of Notre Dame

Assessing silent reading fluency in classroom environments is challenging. This article reports on a method of assessing silent reading using underlining, an approach that solves many problems other silent reading fluency assessment measures face. This method computationally monitors readers' silent reading fluency by the speed they underline words in a text. Traditional silent reading fluency measures were compared with the new underlining methodology. Fourth- and sixth-grade students completed silent reading fluency measures (i.e., moving windows, underlining, and paper-and-pencil reading tasks), along with measures of their oral reading fluency, reading comprehension, and vocabulary knowledge. Strong alternate-form reliability coefficients were found for underlining, which significantly correlated with other measures of silent reading fluency and reading comprehension. Underlining methodology also correlated with common language factors, such as word length and word frequency. Together, these results provide support for the psychometric properties of underlining and suggest that it is a promising alternative method of assessing silent reading fluency. © 2012 Wiley Periodicals, Inc.

Once thought of as a neglected area within the reading literature, oral reading fluency is viewed as a central component of literacy (Kuhn, Schwanenflugel, & Meisinger, 2010). The recognition of reading fluency as a pivotal academic skill has lead to a greater emphasis on reading fluency assessment and instruction (Pikulski & Chard, 2005). Indeed, the ability to read quickly, accurately, and, when reading aloud, with appropriate expression is essential for understanding text (National Reading Panel, 2000). However, the literature has focused almost exclusively on oral reading, leaving silent reading fluency a largely overlooked and understudied phenomenon (Share, 2008).

Oral reading is thought to benefit emergent or struggling readers, as it allows for self-monitoring of progress (Kuhn & Schwanenflugel, 2007), reinforcement of letter–sound correspondence, and the use of both reading and listening comprehension skills to facilitate understanding (Hoover & Gough, 1990; Kuhn & Schwanenflugel, 2007). Once children have become fluent oral readers, they generally transition to silent reading during late elementary school, probably after fourth grade (Prior & Welling, 2001). Skilled child and adult readers rarely read aloud. Yet, literature on reading fluency emphasizes oral reading, creating a mismatch between the everyday literacy activities of many readers and the object of study.

Definitions of reading fluency vary in terms of their relative focus on prosody, reading comprehension, accuracy, and automaticity (see Kuhn et al., 2010, for a review); yet, few theorists differentiate between oral and silent reading. Implicit in the existing literature is the notion that oral and silent reading are essentially the same process. Although the similarities between oral and silent reading cannot be discounted, several researchers (e.g., Share, 2008) have suggested that overarching dependence on oral reading provides an incomplete picture of both reading and reading development. Methodological problems include an overestimation of the importance of phonological variables as well as overstated conclusions about the cognitive processes underlying oral and silent reading. Indeed, eye-tracking research demonstrates that, in skilled readers, the eye tends to be ahead of the

Correspondence to: Elizabeth B. Meisinger, University of Memphis-Psychology, 400 Innovation Drive, 202 Psychology Building, Memphis, TN 38152. E-mail: bmsinger@memphis.edu

voice (e.g., Radach,Schmitten, Glover, & Huestegge, 2009; Rayner & Pollatsek, 1989), suggesting a need to pay closer consideration to oral reading's often ignored counterpart, silent reading. It may be prudent to exercise caution in extrapolating findings from one modality to the other (Share, 2008).

Oral reading fluency is typically assessed by having individuals read aloud from a passage while their reading errors (i.e., miscues) and reading rate are recorded. Curriculum-based measures (CBM) of reading (e.g., the Dynamic Indicators of Basic Early Literacy Skills, Good, Kaminski, & Dill, 2002; AIMSweb, Shinn & Shinn, 2002) are widely used in the school setting (Hintze, Christ, & Methe, 2006; Stecker, Fuchs, & Fuchs, 2005). Children read aloud from each passage for 1 minute, and the number of words read correctly is recorded. With the shift toward the use of the Response to Intervention (RtI) model for the identification of children with learning disabilities, these oral reading probes have increasingly been used to quickly and systematically screen and evaluate large groups of students in classrooms, schools, and districts to make decisions regarding issues such as special education eligibility and diagnoses (Deno, 2003; Stecker & Fuchs, 2000; Wayman, Wallace, Wiley, Ticha, & Espih, 2007).

Several methodologies are currently used to assess various aspects of silent reading. Self-paced reading methods are based on the assumption that a participant will read at a rate that matches the comprehension process, and therefore analysis of the reading time will reveal information about comprehension and the reading process itself (Haberlandt, 1994). For instance, eye-tracking examines where and for how long eye fixations are made, where the eyes regress, and how many times and for how long the reader fixates on individual words (Haberlandt, 1994; Rayner, 1998). Window methods involve a reader successively exposing segments, or windows, of a text on a computer screen. The window may expose a word, phrase, sentence, paragraph, or section of text (Haberlandt, 1994; Rayner, 1998). The participant reveals segments (e.g., word, phrase, or sentence) of the text piece-by-piece by pushing a button. The moving window technique is less precise (Magliano, Graesser, Eymard, Haberlandt, & Gholson, 1993), but both techniques are frequently used in reading comprehension research and yield very similar results (Ferreira & Henderson, 1990; Koornneef & Van Berkum, 2006; Traxler, Pickering, McElree, 2002). Their drawbacks are of a practical nature. Eye-tracking technology only allows one participant to be monitored at a time, involves close monitoring by the administrator, and requires the participant to remain relatively motionless, a difficult task for many children. As a consequence, even though eye-tracking technology yields high-precision reading data, it is expensive and difficult to implement in a classroom environment. Although moving windows is a more feasible approach to use with children, it requires awkward reading behaviors (i.e., pushing a button to reveal each word) and provides more limited reading data (e.g., no information regarding regressions and pauses). Please see Table 1 for a comparative summary of the various silent reading methodologies.

Table 1
Advantages and Disadvantages of Various Silent Reading Assessment Methodologies

	Eye-Tracking	Moving Windows	Paper-and-Pencil	Underlining
Group Administration	−	+	+	+
Relatively Low Cost	−	+	+	+
Intra-Word Reading Data	+	−	−	−
Word Latency Data (i.e., pauses)	+	−	−	+
Regression Data	+	−	−	+
Location of Errors in Text	+	−	−	+
Naturalness of Reading	−	−	+	+

In addition to these self-paced reading methodologies, researchers have also assessed silent reading via paper-and-pencil inventories. Typically, these measures require readers to read silently for a fixed amount of time until an administrator instructs them to circle the last word read (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Paper-and-pencil assessments do not provide the wealth of information available from moving-window methodologies because researchers are unable to glean information about the behaviors that the reader is undergoing while reading (i.e., regressions, pauses, etc.). The self-report nature of this sort of assessment may lead to issues with accuracy and reliability of the measure (Fuchs et al., 2001). However, paper-and-pencil inventories have the advantages that they can be implemented in a classroom setting, allow students to be assessed in groups, and do not deviate from normal reading processes.

In short, our overreliance on oral measures may be due in part to the fact that oral reading is a readily observable behavior. Measuring silent reading fluency in a classroom environment is often impractical. Reading measures commonly used in cognitive science, such as eye-tracking and moving windows (Rayner & Pollatsek, 1989), require individual assessment or button presses on a computer screen. We propose an alternative method that allows assessment of multiple students simultaneously while retaining the level of precision that eye-tracking technologies offer. This method, underlining, takes advantage of younger children often reading while their index finger moves along the words of a sentence. The purpose of the current article is to examine the psychometric properties of underlining as a method of assessing silent reading fluency.

The proposed alternative method of reading assessment takes advantage of the strengths of self-paced reading methodologies and paper-and-pencil inventories, while eliminating their weaknesses (see Table 1). Underlining is a process that combines elements of both paper-and-pencil and eye-tracking methodologies in a less expensive, more child-friendly process and allows for large-scale implementation in the classroom. A passage is presented on the screen of a tablet PC, which has been folded down flat to resemble a piece of paper. Readers are then asked to use the stylus of the tablet computer (PC) to underline the words of the passage "online" with their reading—that is, they are instructed to underline each word as they read it, refrain from underlining if they pause in their reading, and go back and re-underline any words to which they regress within the passage (see Figure 1). Computer software tracks each individual's underlining, providing detailed information about reading rate, pauses, regressions, and other word (and word combination) level features. A first step in the evaluation of this new methodology is to examine its appropriateness for assessing silent reading, an issue that is addressed in this article.

The purpose of the present research was to examine the reliability and validity of this alternative method of assessing silent reading skills, underlining. First, we sought to compare underlining to previously validated measures of silent reading to determine whether it (a) provides similar and consistent results and (b) can replicate common findings in the literature. Second, we compared underlining to measures of reading comprehension to determine its utility as a screener for general reading skill.

Method

Participants

Participants were 37 fourth-grade and 22 sixth-grade students ($N = 59$) attending a public elementary school in the midsouth region of the United States. Fourth-grade participants ranged from 9.34 to 10.57 years of age ($M = 9.90$, $SD = 0.32$); sixth-grade participants ranged from 11.14 to 12.45 years of age ($M = 11.71$, $SD = 0.38$). Students' ethnicity was 56.1% Caucasian, 31.6% African American, 5.3% Asian/Pacific Islander, and 7.0% other or unknown. Half of the participants were girls, and 22% of the students were eligible for free or reduced-priced lunch. Consent forms

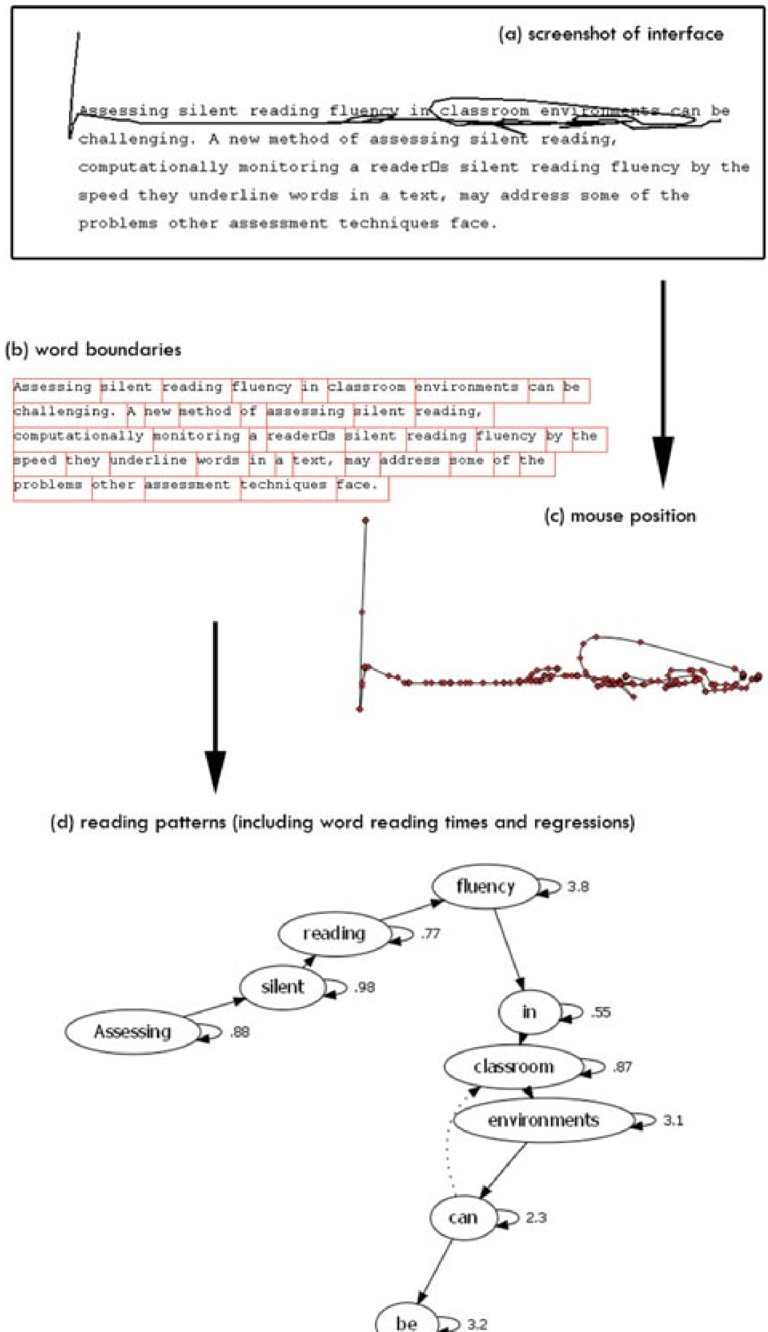

FIGURE 1. Extracting reading patterns from stylus movements. (a) Screen shot of the interface. Stylus positions (c) are aligned with word boundaries (b) to produce reading patterns (d). Numbers on links in (d) represent word reading times in seconds. Dotted links represent regressions.

were provided to the parents of all children attending general education classes in the targeted grades. None were excluded on the basis of reading disability or other special education eligibility. Every student who had both parental consent and child assent were included for participation in the study.

Measures

Reading Passage Selection. Reading passages for the oral and silent reading assessments were selected from the progress monitoring forms of the Dynamic Indicators of Basic Early Literacy Skills, 6th Edition (DIBELS; Good et al., 2002) Oral Reading Fluency subtest. The DIBELS has available 20 passages for each grade level from kindergarten through sixth grade. To ensure that passage difficulty was as close to equivalent as possible across the fluency measures, the readability of each DIBELS passage was measured using three different readability formulas: Dale-Chall, Spache, and Flesch-Kincaid Grade Level. These formulas were selected because they are frequently utilized and are regarded as appropriate for a wide range of ages in the literature (Shapiro, 2004). The average of the three formulas was used to select passages for this study. The readability estimates from each formula were standardized to allow comparisons among the results. The z-scores were then averaged for each passage. Using mean z-scores, nine total expository passages were selected from within the pool of passages for each grade, six for the silent reading measures (i.e., three groups of two passages) and three for the oral reading measure (i.e., one group of three passages). Passages ranged from 304 to 376 ($M = 339.00$, $SD = 23.32$) words in length for the fourth grade and 320 to 376 ($M = 343.33$, $SD = 21.65$) words for the sixth grade.

Oral Reading Fluency. Selected passages from the DIBELS were individually administered to each student to assess proficiency in the oral reading of text. The DIBELS Oral Reading Fluency, 6th Edition (DORF; Good et al., 2002) subtest required each student to read aloud from a story for 1 minute while the examiner recorded misread words (i.e., mispronunciations or substitutions, omissions, and hesitations of more than 3 seconds). After each passage, participants were asked to perform a simple 1-minute retell to help ensure that the passages were actually read. DORF yields a words-read-correctly-per-minute (WCPM) score that is calculated by subtracting the number of word errors from the total number of words read in 60 seconds. In accordance with standard DIBELS procedure, the median WCPM was calculated for participants and was used for comparison for this measure only. Reported reliabilities for DORF ranged from .92 to .97 (Shaw & Shaw, 2002; Tindal, Marston, & Deno, 1983). Correlations with other measures of oral reading and reading comprehension have ranged from .52 to .91 (Shaw & Shaw, 2002; Good & Jefferson, 1998).

Silent Reading Fluency. Children's silent reading fluency was assessed using the underlining, moving windows, and paper-and-pencil procedures described below.

Underlining. The reading passages for this condition were presented in size 14-point Times New Roman font and on a Dell Latitude XT tablet PC. Participants read one practice trial to familiarize themselves with the underlining procedure. Then they completed two trials, each with a different passage. Participants read the passage while underlining the text word-by-word in a smooth motion using the stylus. Participants were instructed to continue to underline online as they read (e.g., if participants regressed, the regression was underlined; if they paused during reading, the underlining paused). During the underlining of each passage, the location of the mouse was recorded at the rate of 10 Hz (i.e., 10 times per second) to track various characteristics of participants' reading (e.g., rate, regressions, pauses, etc.). Software designed specifically for this task was used to track and interpret the mouse coordinates. After each passage, participants were asked to perform a simple 1-minute retell to help ensure that the passages were actually read.

Moving Windows. Passages for the moving windows procedure were presented in 14-point Times New Roman font on the tablet PC described earlier using E-Prime experiment software

(Version 2.0, Psychology Software Tools, Sharpsburg, PA). Participants completed one brief practice trial to familiarize themselves with the moving-window procedure. Then participants completed two trials, each with a specific passage. Passages were presented one word at a time on the tablet PC screen using a self-paced moving-window procedure under control of the participant. Participants were encouraged to move through the passage at their normal reading rate. The reading time was recorded for each word. Again, after each passage, participants were asked to perform a simple 1-minute retell.

Paper-and-Pencil. Following a common procedure, participants were asked to read silently from a passage. At the end of 1 minute, participants were asked to circle the last word read. If participants finished the passage prior to the end of the minute, the exact reading time was recorded (Fuchs et al., 2001). Following each passage, a recall procedure was administered identical to the one described previously.

Reading Comprehension. Comprehension was assessed using the Gates-MacGinitie Reading Test, Fourth Edition (GMRT-4; MacGinitie, MacGinitie, Maria, & Dreyer, 2008) Comprehension subtest. This is a standardized, norm-referenced, group administered test that yields normal curve equivalent scores. Students were asked to silently read a series of 10 passages, accompanied by a series of multiple-choice questions. Participants were given 35 minutes to complete the passages. Test–retest reliability estimates for the GMRT-4 ranged from .83 to .85, and correlations with other tests of reading comprehension ranged from .60 to .62 (MacGinitie et al., 2008).

Comprehension was also assessed by a Reading Maze task from the AIMSweb progress monitoring system (Shinn & Shinn, 2002). The Maze task is a standardized, group-administered cloze task. Students silently read a grade-level narrative fiction passage in which the first sentence was left intact, but thereafter, every seventh word had been replaced by three word choices surrounded by parentheses. One of the three word choices was correct, one was a near distracter (i.e., it was the same word type but did not preserve meaning), and one was a far distracter (i.e., it was not the same word type but did not preserve meaning). Following a short practice passage, each student had 3 minutes to complete the task. Scores were calculated by subtracting the number of incorrect answers from the total number of items attempted. Test–retest reliability estimates of .90 were reported for maze tasks similar to the ones used in this study, and validity estimates ranged from .77 to .85 for students in grades 3 through 5 (Fuchs & Fuchs, 1992).

Vocabulary. The GMRT-4; MacGinitie et al., 2008) Vocabulary subtest is a standardized, norm-referenced, and group administered test of reading vocabulary that yields normal curve equivalent and percentile rank scores. Students are given 20 minutes to answer a series of 45 multiple-choice questions about the meanings of various words. Test-retest reliability estimates for the GMRT-4 ranged from .85 to .91, and validity estimates with other tests of reading vocabulary ranged from .60 to .62 (MacGinitie et al., 2008).

Procedure

Written parental consent and child assent were required for participation in the study. Each child was administered the complete battery of reading fluency, reading comprehension, and vocabulary measures. The assessments of oral and silent reading fluency were administered individually to students in a quiet location in the school and were counterbalanced using a 4-×-4 Latin square. Additionally, reading passages were grouped into three sets of two passages (i.e., groups A, B, and C) and were counterbalanced within each measure to control for reading passage effects. The group-administered assessments (i.e., the GRMT-4 Reading Vocabulary and Reading Comprehension) were conducted with the entire class or with participants collapsed across grades, when appropriate. The graduate students in school psychology who administered the tests received extensive training and

Table 2
Correlations and Descriptive Statistics

Method	1	2	3	4	5	6	7	*M*	*SD*
1. Underlining	–	.56**	.73**	−.62**	−.28*	−.28*	−.50**	0.45	0.15
2. Moving Windows		–	.62**	−.54**	−.18	−.16	−.63**	0.65	0.22
3. Paper-and-Pencil			–	−.72**	−.35**	−.44**	−.47**	0.39	0.15
4. DIBELS Oral Reading				–	−.50**	−.52**	−.54**	0.53	0.16
5. GRMT-4 Vocabulary					–	.61*	.18	62.40	18.34
6. GRMT-4 Comprehension						–	.20	55.91	18.59
7. AIMSweb Maze Task							–	24.80	7.99

Note. GMRT-4 is the Gates-MacGinitie Reading Tests, Fourth Edition. Reading fluency scores are word reading times. GRMT-4 scores are normal curve equivalents.
$^{*}p < .05$, $^{**}p < .01$.

practice. Further, graduate student administrators reached 95% agreement of interrater reliability on DORF prior to the beginning of data collection.

Data Screening. Data were screened to examine for missing data points, outliers, and normalcy. An analysis of the retell data revealed that no participants needed to be dropped due to their failure to appropriately read the passages. Two data points were missing and were estimated using a linear interpolation estimation approach (Tabachnick & Fidell, 2007). Analyses were run with and without the estimated values, and the results were similar, lending confidence in the final dataset. A few outlier scores were identified (i.e., $z > 3.29$) and were decreased to the level of the second highest score within that same measure (Tabachnick & Fidell, 2007). All skewness and kurtosis statistics fell within acceptable limits (i.e., less than 2.0; Tabachnick & Fidell, 2007). For all standardized measures, the normal curve equivalent score was used in analyses. The mean word reading time in seconds (WRT) was then calculated for each participant across the reading fluency measures and was used in subsequent analyses.[1]

Results

Descriptive statistics and correlations for all reading assessments are presented in Table 2. The sample consisted of relatively skilled readers, which is evidenced by sample mean scores somewhat above the national average on the normative assessment techniques (GMRT-4 Comprehension, $M = 55.91$, $SD = 18.59$; GRMT-4 Vocabulary, $M = 62.40$, $SD = 18.59$).

Reliability

Two passages were read for each silent reading fluency measure, making it possible for us to examine the consistency of students' reading performance across passages. The alternate-form reliability coefficients were .88 for paper-and-pencil, .90 for underlining, and .93 for moving windows. These results suggest a high degree of consistency across alternate passages for all three silent reading assessments.

[1] Initially, the average number of words read per minute (WPM) was computed for participants across the four silent reading assessment types. However, the WPM and WRT values for each methodology were found to relate very highly (i.e., $r > .80$). Although results from analyses were similar between each metric, the WRT seemed to be more sensitive to detecting small differences. Therefore, results were presented from the WRT scores.

Reading Rate Across Methodologies

To examine the relation among the four reading fluency methodologies, a series of Pearson product-moment correlation coefficients were computed. All four methodologies were moderately to strongly related with one another (see correlation matrix on the left side of Table 2), suggesting that underlining is comparable to other reading fluency assessments. To ensure that the correlations between reading measures were not due to the effect of age, this and all subsequent analyses were replicated with partial correlations controlling for participants' age in months, and no differences were observed.

A mixed-effects regression analysis was conducted taking into account both participant and item differences, with participants and items as random factors, the three methodologies as fixed factors, and WRT as a dependent variable (Baayen, Davidson, & Bates, 2008; Locker, Hoffman, & Bovaird, 2007; Richter, 2006). The mixed-effects technique allows for by-subject and by-passage analyses to be conducted concurrently as opposed to separately, as in a traditional analysis of variance or regression model. The major advantage of this methodology is that results can be generalized not only across participants, but also across stimuli (i.e., reading passages). The model was fitted using the restricted maximum likelihood estimate for the continuous variable (WRT). A significant difference between the methodologies was found, F (3, 397) = 94.5, MSE = 1.41, $p < .001$.Tukey post-hoc analyses at $p < .05$ yielded the following pattern in reading rate: moving windows > oral reading > underlining > paper-and-pencil (see the right side of Table 1). Notably, the reading rate yielded by underlining was more similar to that of the ecologically valid paper-and-pencil measure than to the moving windows method. The paper-and-pencil task is most similar to natural silent reading (though it reveals little about the reading process itself), whereas the moving windows technique is somewhat awkward and unnatural. Therefore, the similarity between the reading rate yielded by the underlining and the paper-and-pencil assessments supports the validity of underlining as a measure of silent reading fluency.

Relation to Comprehension and Vocabulary Measures

To determine the strength of the relations between the various reading fluency assessments and the reading comprehension and vocabulary measures, a series of Pearson product-moment correlations were computed. An examination of the distribution statistics revealed concerns with restriction or expansion of range. Therefore, correlations were corrected using a formula that utilizes the known population standard deviation for the normed GRMT-4 Comprehension and Vocabulary measures to provide a more accurate representation of the relations between variables (Cohen, Cohen, West, & Aiken, 2003; Gulliksen, 1987). The corrected and uncorrected correlation coefficients were generally comparable; therefore, only the corrected coefficients are presented in Table 2. Importantly, results indicated that underlining was consistently related to children's comprehension and vocabulary skills (see Columns 5-7, Row 1, in Table 2). Likewise, the paper-and-pencil and DORF tasks were also significantly related to reading comprehension and vocabulary. However, when children's silent reading fluency was measured via the moving windows method, a relation was found only with the Maze task. Thus, the Maze task correlated more strongly with the reading measures than did the other two comprehension tasks. Interestingly, the Maze task did not correlate significantly with either of the standardized comprehension and vocabulary tasks, bringing into question whether the Maze was measuring more than reading rate. Still, the fluency measures were moderately to strongly related to the Maze task, suggesting that an important association exists.

Replication of Common Findings

To ensure that the underlining technology is capable of replicating common patterns found in the literature, some additional analyses were conducted. Specifically, the accuracy of the derived reading times was verified by determining that reading times were positively correlated with word length and negatively correlated with word frequency (Rayner & Pollatsek, 1989). The analyses proceeded by randomly selecting 100 words from the passages and computing the mean reading time for each word across participants. Word length was computed as the number of characters and was positively correlated with mean reading times for both fourth-grade students ($r = .53$, $p < .001$) and sixth-grade students ($r = .56$, $p < .001$).

Word frequencies were obtained from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). Inverse weights were computed from the word frequencies. Specifically, each word was weighted on a scale of from 0.0 to 1.0, relative to its inverse frequency in the English language. As a consequence, higher frequency words, such as closed-class function words (e.g., *and*, *but*, *a*, *the*) have comparatively low weights, whereas lower frequency words (e.g., *cluster*, *oxbow*) have higher weights. The obvious prediction is that high-frequency words (low weights) should have shorter reading times than low-frequency words (high weights). Therefore, we expected a positive correlation between reading times and inverse weighted frequencies. Consistent with the previous literature (Balota & Chumbley, 1984; Rayner & Pollatsek, 1989), a positive correlation was observed between mean reading times and inverse weighted frequency for fourth-grade students ($r = .25$, $p < .05$) and sixth-grade students ($r = .24$, $p < .05$). These findings confirm our prediction that reading times were longer for lengthier words as well as for less frequent content words (Balota & Chumbley, 1984; Rayner & Pollatsek, 1989).

Discussion

The aim of this study was to examine the reliability and validity of underlining, an alternative method of assessing silent reading fluency. Underlining attempts to address many of the drawbacks that are present in current methods of silent reading assessment. Our results provide strong support for the use of underlining as a means of assessing silent reading fluency. In particular, underlining may represent a useful tool for researchers who seek to assess the effectiveness of fluency-oriented reading interventions or to elucidate developmental trends in reading fluency.

To evaluate the appropriateness of underlining as a measure of silent reading fluency, we first examined the consistency of student performance for each method across alternate passages. Alternate-form reliability estimates for underlining fell above acceptable limits (i.e., $r > .80$). Next, we compared its results with those garnered from other commonly used and previously validated measures of silent reading fluency. Strong interrelations were found between underlining and the oral and silent reading assessments. Using the underlining methodology, we were also able to replicate common findings in the literature (i.e., correlations in the expected direction between WRTs and word length and frequency). Significant correlations with external fluency variables and the replication of common response processes provide evidence for the validity of underlining. These findings, along with strong reliability coefficients, provide evidence supporting the psychometric properties of underlining and suggest it is a good alternative to existing measures.

As would be expected for a sound measure of silent reading fluency, the underlining task consistently and significantly correlated with the standardized, norm-referenced reading comprehension and vocabulary tasks. However, underlining demonstrated a more robust association with the maze task, a commonly used screener for reading comprehension within RtI models (Deno et al., 2009), suggesting that underlining may provide similar information to measures currently used in the schools for these purposes. Yet, the maze task did not correlate significantly with either of the normed comprehension measures used in this study. Previous research found significant positive correlations between (oral and silent) reading fluency and reading comprehension, as well as between maze tasks and reading comprehension; however, reading fluency generally appears to be a stronger predictor of reading comprehension than maze tasks (Ardoin et al., 2004; Denton et al., 2011; Jenkins & Jewell, 1993). It may be that the maze procedure represents a relatively surface-level comprehension task, whereas the standardized measure used in this study (i.e., GRMT-4 Comprehension subtest) involves more in-depth passage reading, requiring both literal and inferential responses from the reader. Importantly, the relation between underlining and comprehension is comparable and at times superior to the other computerized task, moving windows. It should be noted that the observed correlations between underlining and comprehension in this study are roughly equivalent to those reported in the literature for the interrelations between various reading comprehension measures (cf. Keenan, Betjemann, & Olson, 2008).

Educational Implications

A need exists for a sound silent reading fluency measure that is appropriate for school settings. Oral reading fluency assessments (i.e., CBM oral reading probes) and interventions are prevalent in today's schools (Kuhn et al., 2010). Yet, beyond the early elementary schools years, the completion of everyday academic tasks requires proficient silent reading skills. If children struggle to read assignments in content area textbooks or to accurately read test items in the time allotted, they are likely to fall behind their peers and experience poor academic achievement (Stanovich, 1985). It seems essential to evaluate whether, as a result of a reading fluency intervention, students are better able to silently read grade-level connected text. In classrooms outfitted with several tablet PCs, data could be collected simultaneously from multiple students via a wireless router throughout the school day with minimal disruption to instruction. After mastering the procedure, children could complete underlining as part of their regular routine with little supervision, allowing for the continuous collection of progress monitoring data without the expense of additional personnel. In lieu of a retell, brief comprehension questions could augment underlining to ensure that children read for understanding. Given the prevalence of the RtI framework in the schools, the identification of appropriate progress monitoring techniques is vital for school psychologists and other school professionals.

Due to the dearth of research on this topic, the exact nature of the relation between oral and silent reading fluency is unclear. In particular, the degree to which oral and silent reading fluency involve overlapping processes (e.g., phonological processing) warrants investigation (Share, 2008). Also, the age at which children transition from effective oral to silent reading, as well as how this shift is affected by reading ability, is not well understood. The impracticality (i.e., eye-tracking), awkwardness (i.e., moving windows), or dependence on self-report (i.e., paper-and-pencil) of silent reading measures may have hindered growth in this important area of study. Knowledge of basic developmental trends in reading fluency may inform assessment practices. At present, it is not clear whether children referred to a school psychologist due to suspected reading disability should be administered measures of oral reading fluency, silent reading fluency, both, or none, depending on their ages. It is our hope that the underlining method will provide a feasible means for addressing these important pedagogical and theoretical issues.

Limitations and Future Directions

Several limitations of this work warrant discussion. First, the sample is relatively small, and the patterns warrant replication with a larger sample. Additional evidence regarding the consistency of underlining across administration sessions (i.e., test–retest reliability), as well as predictive validity evidence, would further bolster support for this technique. However, it is notable that the topic of the study is novel and that the results showed support for underlining as a measure of silent reading fluency. Nevertheless, a more detailed examination of underlining is needed across grade levels, perhaps extending into the middle-school years. Second, the sample consisted of relatively skilled readers. Future investigations should utilize a child sample that includes more diverse reading abilities to determine how variance in this regard impacts the results. Similarly, the role of text difficulty on the results garnered from the underlining methodology needs to be examined. Consistent with common progress monitoring practices in the schools, all of the children read grade-level text, precluding the examination of text difficulty. Lastly, the interaction between text difficulty level and reading skills on the appropriateness of silent reading assessment techniques in elementary-school children warrants investigation.

Conclusion

Overall, it appears that underlining is a viable tool for assessing silent reading fluency. Underlining is very adaptable, allowing it to be as brief or as lengthy as the assessor chooses. It can be used with any type of text and allows for a greater analysis of the reading characteristics than do traditional paper-and-pencil techniques (i.e., allows for examination of regressions and pauses and the differences in reading rate across various parts of the passage), which could inform remediation studies involving struggling readers. It also provides a greater assurance that the individual is indeed reading the passage and combats the issue of relying on the self-report of the examinee to determine silent reading rate (as is done with paper-and-pencils methods), which could provide falsely inflated scores, especially in struggling readers. Underlining may be particularly useful for researchers interested in collecting periodic progress monitoring data for the purpose of evaluating the effectiveness of reading interventions. In a classroom outfitted with tablet PCs (or even a regular PC or laptop using a mouse), data regarding children's reading could be collected throughout the day via a wireless router without disrupting instruction. Lastly, its brevity and adaptability may especially lend it to the use of CBM to quickly and frequently assess students' reading progress.

References

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Restar, J. L., et al. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. School Psychology Review, 33(2), 218–233.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59, 390–412.

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (CD-ROM). Philadelphia: University of Pennsylvania.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Verbal Learning and Verbal Behavior, 22, 88–104.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Deno, S. L. (2003). Developments in curriculum-based measurement. Remedial and Special Education, 37, 184–192.

Deno, S., Keschly, A. L., Lembke, E. S., Magnusson, D., Callender, S. A., Windram, H., et al. (2009). Developing a school-wide progress monitoring system. Psychology in the Schools, 46, 44–55.

Denton, C. A. Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S. Cirino, P. T. et al. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. Scientific Studies of Reading, 15(2), 109–135.

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. Journal of Experimental Psychology: Learning, Memory, & Cognition, 16, 555–568.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. School Psychology Review, 21, 45–58.

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

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), Advanced applications of curriculum-based measurement (pp. 61–88). New York: Guilford Press.

Good, R. H., Kaminski, R. A., & Dill, S. (2002). DIBELS Oral Reading Fluency. In R. H. Good & R. A. Kaminski (Eds.), Dynamic Indicators of Basic Early Literacy Skills (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Gulliksen, H. (1987). Theory of mental tests. Hillsdale, NJ: Lawrence Erlbaum.

Haberlandt, K. (1994) Methods in reading research. In M. A. Gernsbacher (Ed.), Handbook of psycholinguistics (pp. 1–32). San Diego, CA: Academic Press.

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. Psychology in the Schools, 43, 45–56.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. Reading and Writing, 2, 127–160.

Jenkins, J. R. & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. Exceptional Children, 59, 421–432.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. Scientific Studies of Reading, 12, 281–300.

Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. Journal of Memory and Language, 54, 445–465.

Kuhn, M. R., & Schwanenflugel, P. J. (2007). Fluency in the classroom. New York: Guilford Press.

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. Reading Research Quarterly, 45, 232–253.

Locker, L., Jr., Hoffman, L., & Bovaird, J. A. (2007). The use of multilevel modeling in the analysis of psycholinguistic data. Behavior Research Methods, 39, 723–730.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2008). Gates-MacGinitie Reading Tests, Fourth Edition Technical Report Supplement Forms S and T. Rolling Meadows, IL: Riverside Publishing.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2007). Gates-MacGinitie Reading Tests, Fourth Edition. Rolling Meadows, IL: Riverside Publishing.

Magliano, J. P., Graesser, A. C., Eymard, L. A., Haberlandt, K., & Gholson, B. (1993). Locus of interpretive and inference processes during text comprehension: A comparison of gaze durations and word reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 704–709.

National Reading Panel. (2000). Report of the national reading panel. Washington, DC: Author.

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. The Reading Teacher, 58(6), 510–519.

Prior, S. M., & Welling, K. A., (2001). "Read in your head": A Vygotskian analysis of the transition for oral to silent reading. Reading Psychology, 22, 1–15.

Radach, R., Schmitten, C., Glover, L., & Huestegge, L. (2009). How children read for comprehension: Eye movements in developing readers. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), Beyond decoding: The behavioral and biological foundations of reading comprehension (pp. 75–106). New York: Guilford Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372–422.

Rayner, K., & Pollatsek, A. (1989). The psychology of reading. New York: Routledge.

Richter, T. (2006). What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. Discourse Processes, 41, 221–250.

Shapiro, E. S. (2004). Academic skills problems: Direct assessment and intervention (3rd ed.). New York: Guilford Press.

Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. Psychological Bulletin, 134, 584–615.

Shaw, R., & Shaw, D. (2002). DIBELS Oral Reading Fluency-Based Indicators of Third Grade Reading Skills for Colorado State Assessment Program (CSAP) (Technical report). Eugene, OR: University of Oregon.

Shinn, M. R., & Shinn, M. M. (2002). AIMSweb training workbook: Administration and scoring of reading maze for use in general outcome measurement. Eden Prairie, MN: Edformation.

Stanovich, K. E. (1985). Mathew effects in reading: Some consequences of individual differences in the acquisition of literacy. Reading Research Quarterly, 21, 360–407.

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum based measurement: The importance of individual progress monitoring. Learning Disabilities Research and Practice, 15, 128–135.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. Psychology in the Schools, 42, 795–819.

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.), Boston, MA: Pearson.

Tindal, G., Marston, D., & Deno, S. (1983). The reliability of direct and repeated measurement (Research Report No. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. Journal of Memory and Language, 47, 530–547.

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