Peer-Assisted Assessment in Reading: Two Exploratory Studies

Paul J. Riccomini,[1] Sheri Berkeley,[2] Allison Neally,[3] Christina Stagliano,[3] Leigh Ann Kurz,[2] and Frederick Brigham[2]

Abstract

Oral reading fluency (ORF) assessments measure how quickly and accurately students read within one minute. They are widely used at the elementary level; however, due to the typical structure and class sizes in middle schools, such individualized assessments are less feasible and, therefore, less frequently used. Two exploratory studies investigated potential methods for efficiently administering ORF measures at the middle school level by utilization of peers: Peer-Assisted Assessment in Reading (PAAR). Findings from both studies showed that after a short training, students were highly accurate in identifying words read in one minute regardless of grade or instructional level, but they were less consistent in correctly identifying miscues. Comparisons between student and student teacher accuracy were also made. Implications for practice are discussed.

Keywords: peer assisted, oral reading fluency, middle school, preservice teachers

Reading and assessment are two of the more widely discussed aspects of current American educational systems. For example, assessments are embedded in major federal regulations both in general (i.e., Every Student Succeeds Act, 2015) and special education (i.e., Individuals With Disabilities Education Improvement Act, 2004). Further, the American Recovery and Reinvestment Act (2009) included additional emphasis on the use of assessment data to improve educational outcomes by offering funds to states that close achievement gaps between the general population and students with disabilities, students from minority backgrounds, low-income students, and students with limited proficiency in the English language. The emphasis on verifying academic performance through assessments is also clear in the Common Core State Standards that have been adopted by a wide majority of states (Calkins, Ehrenworth, & Lehman, 2012; Kendall, Ryan, Alpert, Richarson, & Schwols, 2012).

The initial assessments created in the wake of regulations were most often static end-of-course assessments to validate progress and educational accountability. Such assessments are summative in nature and are of little value in adjusting instructional procedures for individuals who are enrolled in a course (Brigham, Tochterman, & Brigham, 2000). However, a wider array of sensitive measures that can track progress across the school year are available now to educators and are becoming prominent in professional practice in schools (Brigham, Berkeley, & Walker, 2012). For example, progress monitoring is increasingly employed by schools to detect and address educational problems at an early point (Berkeley & Riccomini, 2017; Deno, 2003; Fuchs, Fuchs, & Compton, 2012; Petscher, Cummings, Biancarosa, & Fien, 2013; Santi & Vaughn, 2007).

Among the common elements of progress monitoring procedures is the use of materials that are drawn directly from the instructional materials to be used with students, or at least are very similar to those that students will encounter in their studies (Venn, 2013). Using the same or similar materials for both instruction and assessment has the benefit of reduc-

[1]The Pennsylvania State University; [2]George Mason University; [3]The University of Georgia

Riccomini,P. J., Berkeley, S., Neally, A., Stagliano, C., Kurz, L., & Brigham F. (2019). Peer-Assisted Assessment in Reading: Two Exploratory Studies. *International Journal for Research in Learning Disabilities, 4*(1), 12-26. https://doi.org/10.28987/ijrld.4.1.12

ing the amount of inference required to interpret the assessment data, thereby enhancing their predictive validity (Lembke, Hampton, & Hendricker, 2013) as well as their validity in guiding interventions, also known as treatment utility (Hayes, Nelson, & Jarrett, 1987). Hayes et al. (1987) defined treatment utility as simply "the degree to which assessment is shown to contribute to beneficial treatment outcome" (p. 963). Roach (2008) described the amount of inference required to employ assessment data in useful interventions as the "X factor," in which the connections between assessments and instructional actions are often tenuous and poorly understood. Thus, reducing the conceptual distance between measures and the actions to be based upon them is likely to minimize error and response time as well as make the process of assessment more transparent.

Several variables have been identified as useful targets for screening students for risk of failure and for monitoring general development within a performance domain. Most of these are fluency variables where speed and accuracy of task completion are important (Schwanenflugel & Ruston, 2008).

Reading Assessment and Middle School

The goal of reading is comprehension. Reading involves a union of decoding, fluency, and comprehension that is integrated and automatized in effective readers so that they are able to devote their full attention to the meaning of the text (Soriano, Miranda, Soriano, Nievas, & Felix, 2011). In contrast, readers who are not fluent in reading understand less of what they read, are less motivated to practice, and struggle more in learning academic content (Kuhn & Stahl, 2003; Meisinger, Bloom, & Hynd, 2010). While adequate fluency does not ensure good comprehension of text, poor fluency virtually ensures that the reader will have a great deal of difficulty in grasping the meaning of what he or she reads.

It is well established that oral word reading fluency is a strong predictor of overall reading competence (e.g., Fuchs, Fuchs, Hosp, & Jenkins, 2001; Harniss, Caros, & Gersten, 2007; Jenkins, 2009; Stecker, Fuchs, & Fuchs, 2005) and that a key relationship exists between fluency and comprehension (Hosp & Fuchs, 2005; Neddenriep, Fritz, & Carrier, 2011; Petscher et al., 2013). It is important to note that the relationship between ORF and comprehension decreases with reading development (e.g., Denton et al., 2011; Jenkins & Jewell, 1993; Silberglitt, Burns, Madyun, & Lail, 2006). However, although the relationship is lower than among elementary children, ORF is still widely used as a screening and progress monitoring measure with students in middle school.

One simple and effective method of assessing fluency is to compute the correct number of words read aloud per minute (CWPM). To conduct this assessment, teachers listen to an individual student read for one minute, note the number of words read and the number of errors, and compute the student's reading rate (Hasbrouck & Tindal, 2006). The structure of elementary schools makes individual assessment more practical than at the secondary levels. At the elementary level, teachers have 20 to 30 students in the same class for much of the school day. In addition, schools often have extended language arts blocks where teachers and/or reading specialists work with students in small groups or individually on a regular basis. Within this context, monitoring of student progress in reading, including fluency development, is a part of typical practice that emphasizes instruction in basic reading.

Middle and high schools, on the other hand, are typically organized by specific disciplinary classes (e.g., science, social studies, English, math) that become increasingly specialized by high school (e.g., biology, chemistry, physics) (Siskin, 1994; Troia, 2006). This organizational structure creates time constraints as teachers are responsible for a larger number of students (e.g., 30 students in up to 8 classes) with whom they have contact for a shorter amount of instructional time (e.g., 45-60 minute instructional periods), which, in turn, limits individualized assessment opportunities.

Another limiting factor at the middle school level is the amount of content that teachers are required to cover within the curriculum, including English language arts classes, where basic reading instruction is no longer a focus. The amount of time that teachers need in order to cover the required curriculum limits the class time available for individualized assessment or instruction. Therefore, individually administered types of assessments at the middle school level are generally not feasible (Berkeley & Riccomini, 2017). For this reason, teachers often opt for group-administered alternate assessments of comprehension (e.g., MAZE; Hosp & Hosp, 2003) rather than individual assessments of oral reading ability (Barth, Stuebing, Fletcher, Denton, Vaughn, & Francis, 2014; Espin, Wallace, Lembke, Campbell, & Long, 2010).

This is highly unfortunate, because as students progress in school, reading fluency becomes more critical in content-area classes. That is, students are

expected to read greater amounts of more difficult material in less time. Because the focus of instruction is on subject content, rather than basic reading skills, adolescents with reading problems are often overlooked and struggle with these reading tasks (Mercer, Campbell, Miller, Mercer, & Lane, 2000). Therefore, teachers would be well advised to consider the reading fluency levels of their students when selecting instructional texts and contemplating the provision of instructional supports. Doing so means that teachers need to find alternative ways to assess students' reading fluency.

Alternative Administration

While ORF measures may not be practical for ongoing progress monitoring, they may be feasible as a screening measure that can help inform instruction both classwide and for individual students. The information gained from a classwide screening of reading fluency can assist a teacher in selecting instructional materials that match the skill levels of the class (Hosp, Hosp, & Howell, 2016) as well as selecting instructional methods/approaches to build background knowledge and vocabulary that may help students compensate for weaknesses in basic reading skills (Espin & Deno, 1995; Espin, Shin, & Busch, 2005). The information gained can also assist a teacher in quickly identifying students who do not have satisfactory basic reading skills, making modifications or accommodations based on this information, targeting additional supports for the student, and monitoring the progress of the student more closely throughout the year (Hosp et al., 2016; Wayman, Wallace, Wiley, Ticha, & Espin, 2007).

A possible solution to the logistical dilemmas and time constraints associated with individually administered assessments at the secondary level may lie in findings from studies on classwide peer tutoring (CWPT). CWPT is an instructional approach where students are taught by their peers (Greenwood, Maheady, & Delquadri, 2002). Unlike other approaches that utilize peers, such as cooperative learning, CWPT employs a large amount of structure, with both tutor and tutee being trained in tutoring procedures and supervised by classroom teachers (Maheady, Harper, & Mallette, 2003).

The benefits of peer tutoring include the ability of students to (a) learn more in less time, (b) improve oral reading rates, and (c) increase academic responding (Morano & Riccomini, 2017). It has been successfully implemented at the secondary level in the areas of mathematics (e.g., Calhoon & Fuchs, 2003), social studies (e.g., Marshak, Mastropieri, & Scruggs, 2011), and English language arts (Mastropieri et al., 2001). Further, students have been successfully paired to work together to improve reading (e.g., Fuchs, Fuchs, & Kazdan, 1999; Harris, Marchand-Martella, & Martella, 2000; Spencer, Scruggs, & Mastropieri, 2003). Because students can be taught to work together to teach each other content and reading skills, it seems probable that students could also be taught to recognize and note errors when listening to each other read.

The Current Study

Peer-Assisted Assessment in Reading (PAAR) has the potential to help middle school teachers screen entire classes of students in order to quickly and efficiently identify students who are not fluent readers and are likely to struggle. Using PAAR, students' ability to read text is generally described using three categories with the following criteria: independent (≥97% accuracy), instructional (90-96% accuracy), and frustrational (<90% accuracy) (see Gillett, Temple, Temple, & Crawford, 2017).

In order to investigate whether PAAR is viable for middle school classrooms with a wide range of reading abilities, two exploratory studies investigated whether students who read below grade level could accurately score a wide range of readers, including more proficient readers (i.e., levels above their instructional level). Specifically, the following research questions were investigated:

1. Can struggling readers accurately determine the number of words read and miscues made during a one-minute ORF assessment by (a) an average reader (instructional level), (b) an above-average reader (independent level), and (c) a below-average reader (frustrational level)?
2. Does student performance differ when reading text that is (a) on grade level, (b) one year below, (c) two years below, or (d) three years below?

In addition to the above questions, Study 2 also investigated how accurate students were compared to experienced and preservice teachers; specifically, Is student scoring of words read and miscues made comparable to that of (a) master's-level student teachers, and (b) undergraduate student teachers?

General Method

Overview

To explore the plausibility of using PAAR as a fluency screening administration procedure, researchers developed audio passages read by a range of readers (independent, instructional, and frustrational) with a series of passages (on grade level, and one, two, and three years below). Then participants were trained in ORF scoring. Finally, students listened to and scored the pre-recorded audio passages.

Audio passages. The audio stories used for the practice activities were developed by the two lead researchers. For each grade level (3 through 8), three comparable passages from the Reading Fluency Monitor (RFM; Read Naturally, 2008a) were selected and designated as *frustration, instructional,* or *independent*. All passages were expository in nature (see Table 1 for more information about the passages selected). Then a corresponding rate (number of words read) and errors (number of miscues) were determined for each passage. The planned errors included mispronunciations, omissions, and hesitations, consistent with scoring conventions of oral reading fluency (see Gillet et al., 2017).

Once the errors were included, the lead researcher read aloud the story, making the predetermined errors and stopping at the predetermined number of words read while being recorded by a standard digital microphone. Each digital audio clip included the stating of the directions, the reading aloud of the story with predetermined miscues, and a beeping sound to indicate the beginning and ending of the one-minute time frame. Links to these audio files were embedded into a PowerPoint slide for easy access.

Measure

Test of scoring proficiency. The Test of Scoring Proficiency captured participants' accuracy in scoring the prerecorded audio passages from the RFM. For each passage, participants notated the last word read in one minute and reading miscues, and then calculated correct words per minute (CWPM). Two independent researchers counted the number of words read and miscues that were notated by participants, computed CWPM, and reconciled any differences from 95% (mostly due to illegible writing) to 100% agreement. These scores were used in the analysis.

Procedures

Participant training. One of the researchers provided training to all participants. At the time of the study, the trainer had been providing professional development on oral reading fluency to practicing teachers for seven years. The trainer held a doctoral degree in special education and had previously taught in middle school. The trainer also developed the training materials and activities for the project based on previous work (Riccomini & Stecker, 2005).

After a brief overview of oral reading fluency, the trainer provided a model for each scoring convention with an example that was displayed to participants

Table 1
Passage Technical Information

Grade-Level Passages	Length of Passage Range	Benchmark Correlation Range[a]	Range of Difficulty[b]
3	152 to 202	.92 to .97	.4 to .7
4	171 to 210	.95 to .96	-1.2 to 1.2
5	237 to 254	.92 to .95	-4.1 to 3.9
6	217 to 249	.89 to .95	-5.4 to -.5
7	231 to 247	.93 to .95	-1.3 to -3.9
8	228 to 311	.91 to .95	-4.7 to -3.07

[a]Benchmark correlations are a measure of validity with correlations of .8 or higher.

[b]Range of difficulty indicates median words correct per minute of norming sample compared to medium benchmark. Negative scores are more difficult than the benchmark and positive scores are less difficult than the benchmark. Information available from: https://www.readnaturally.com/knowledgebase/documents-and-resources/25/296

using PowerPoint. Following the model, the trainer reviewed each scoring rule, and then the participants were given an opportunity to practice scoring single sentences that contained the target error. The participants' performance was closely monitored, and any errors were re-explained until corrected. Students had to demonstrate a 90% scoring accuracy before proceeding to online practice activities.

Scoring. Participants were given pencils and folders containing sets of three passages (frustrational, instructional, independent) for each grade level: on the student's grade level as well as one, two, and three years below that grade level (N = 12). Passages were randomly ordered by reading level and organized in increasing difficulty from three years below to grade level. When scoring passages, participants removed the first passage from the folder, placed the headphones over their ears, clicked on the audio file that matched the passage, and completed the scoring. This process was repeated until all passages were scored.

Data Analysis

Two procedures were used to compare participant scores of fluency passages to the accurate scoring value. First, a series of one-sample *t*-tests were used to help determine whether the population mean (participant scoring) was equal to our hypothesized valued (accurate scoring). When employing multiple *t*-tests, an adjustment of alpha (e.g., Bonferroni) is typically employed to correct for a familywise error rate. However, for the purpose of this study, an uncorrected alpha of .05 is a more rigorous standard, and thus was employed. In addition, because scores on several passages did not follow a normal distribution, a Wilcoxon Signed-Rank Test was used to verify the results from the *t*-tests. Finally, 95% confidence intervals (CI) were determined and visually analyzed to evaluate whether the upper and lower ranges were consistent with acceptable error ranges when administering and scoring ORF assessments.

Experiment 1

Method

The general methods described above were used for this study. A description of the participants and specific training procedures specific to Experiment 1 follows.

Participants. The participating middle school was located in a small rural district in the southeastern United States; it had an enrollment of 1,000 students in grades six, seven, and eight. The student population was 49% male, and student ethnicities were: 86% Caucasian, 10% African-American, 3% Hispanic, and less than 1% Asian/Pacific Islander. The school was categorized as Title I, where more than 50% of students received free or reduced-price lunch. In the 2005-2006 and 2006-2007 school years, the school did not make "Adequate Yearly Progress" and had a "Needs Improvement" status. Students were selected by the school for a remedial reading class based on the results of the prior year's high-stakes tests in reading and language arts. The required reading class replaced one of the students' elective classes.

Student assent and parental consent were obtained for 88 students enrolled in the reading class; however, nine students were absent during one or more of the testing sessions and, therefore, were not included in the final data analysis. The resulting number of subjects was 79 students (sixth grade = 31, seventh grade = 22, eighth grade = 26) – 32 boys and 47 girls. Subjects were primarily Caucasian (75.9%), followed by African American (12.7%), Hispanic (7.6%), and mixed ethnicities (3.8%). No students with identified disabilities in reading participated in the study.

Prior to the study, a series of assessments was given to gain a comprehensive picture of participants' reading levels. All students scored below the 1st percentile in total word reading (phonetic decoding and sight word recognition) on the Test of Word Reading Efficiency (TOWRE; Torgesen, Rashotte, & Wagner, 2012). In addition, students at all grade levels read fewer correct words per minute (CWPM) on grade-level passages of the Reading Fluency Benchmark Assessor (RFBA; Read Naturally, 2008b) than expected for their respective grade when compared to national norms (see Hasbrouck & Tindal, 2006). Specifically, sixth graders read an average of 105.53 CWPM vs. the expected rate of 127 CWPM; seventh graders read an average of 83.27 CWPM vs. the expected rate of 128 CWPM; and eighth graders read an average of 66.83 CWPM vs. the expected rate of 133 CWPM. To provide a more completed description of the participants, additional achievement data for the Gates-MacGinitie Reading Tests (GMRT; MacGinitie, MacGinitie, Maria, & Dreyer, 2000) and student lexiles are presented in Table 2.

Training procedures. All participants received a 45- to 60-minute training session on the parame-

Table 2
Study 1 & 2 Student Achievement Data by Grade Level

	RFBA CWPM *M* (range)	GMRT %tile *M* (range)			Lexile *M* (range)
		Vocabulary Subtest	Comprehension Subtest	Total Score	
Study 1					
6th Grade (*N* = 28)	94.51 (13 to 145)	21.00 (2 to 66)	26.58 (4 to 62)	25.48 (7 to 57)	735.00 (525 to 935)
7th Grade (*N* = 12)	87.67 (49 to 146)	21.25 (1 to 72)	16.42 (1 to 54)	18.71 (1 to 50)	702.92 (525 to 935)
8th Grade (*N* = 11)	73.24 (44 to 118)	10.27 (1 to 34)	13.45 (1 to 31)	10.64 (2 to 32)	741.36 (560 to 875)
Study 2					
6th Grade (*N* = 16)	98.25 (48 to 136)	26.68 (1 to 37)	35.37 (15 to 60)	32.11 (15 to 55)	763.42 (650 to 885)
7th Grade (*N* = 13)	84.78 (23 to 130)	11.92 (1 to 23)	17.83 (6 to 34)	13.25 (1 to 27)	728.33 (630 to 820)
8th Grade (*N* = 21)	67.45 (35 to 105)	17.18 (1 to 50)	16.45 (3 to 42)	15.73 (2 to 48)	757.73 (630 to 910)

Note. RFBA CWPM: Reading Fluency Benchmark Assessor Correct Words Per Minute; GMRT: Gates-MacGinitie Reading Tests.

ters and scoring conventions for oral reading fluency (described under General Procedures). Participants were trained using single sentences and not complete paragraphs similar in length and frequency of miscues to those used to measure their scoring. The scoring conventions included in the training were mispronounced words, omissions, word reversals, hesitations, and self-corrections.

Results

Student scoring. As shown in Table 3, findings from analysis of student accuracy in scoring *words read* in one minute produced insufficient evidence to conclude that student scoring errors were significantly different from zero (i.e., accurate scoring). Further, statistical analysis could not be conducted on 12 of 36 passages (33%) because all students scored the passage accurately. Visual inspection of confidence intervals showed that upper and lower ranges were within acceptable ranges on all passages.

Statistically significant differences between student scoring of reading *miscues* and zero (i.e., accurate scoring) for each of the passages were found. However, visual inspection of confidence intervals showed that upper and lower ranges were only marginally outside acceptable ranges on all passages (see Table 4).

Counting and computation accuracy. Counting and subtraction errors were tallied to determine accuracy (see Table 5). Overall, students had high levels of accuracy counting notated words read (*M* = 88%), counting notated reading miscues (*M* = 99%), and computing CWPM (*M* = 88%).

Experiment 2

Method

The general methods mentioned above were also utilized for Study 2; however, this study extended the work of Study 1, as follows. First, the training was streamlined so that the critical content could be addressed within 20 minutes. In addition, students were required to pass a "certification test" with 90% accuracy before proceeding to scoring the prerecorded audio passages. Finally, in addition to comparing student performance to accurate scoring, student scoring performance was compared to the performance of master's-level and undergraduate student teachers who were trained to score an ORF as part of their coursework. A description of participants and revised training procedures follows.

Participants. Students were recruited from the same school as in Study 1 during the following school year. No students participated in both Study 1 and Study 2, so participants in Study 2 represented an independent sample. Student teachers were recruited from a university teacher preparation program.

***Students*.** Participating students consisted of 21 sixth graders, 13 seventh graders, and 16 eighth

Table 3
Study 1: Student Identification of Words Read

	8th Grade (*N* = 11)		7th Grade (*N* = 12)		6th Grade (*N* = 28)	
	M (SD)	95% CI	*M (SD)*	95% CI	*M (SD)*	95% CI
On Grade Level						
Independent	0.00 (0.00)	[0.00, 0.00]	0.17 (0.58)	[-0.20, 0.53]	2.14 (11.14)	[-2.18, 6.46]
Instructional	3.09 (10.25)	[-3.80, 9.98]	-1.25 (4.33)	[-4.00, 1.50]	0.00 (0.00)	[0.00, 0.00]
Frustrational	0.00 (0.00)	[0.00, 0.00]	-1.00 (3.46)	[-3.20, 1.20]	-1.11 (6.26)	[-3.53, 1.32]
One Level Below						
Independent	0.55 (1.29)	[-0.32, 1.41]	0.00 (0.00)	[0.00, 0.00]	-1.43 (7.56)	[-4.36, 1.50]
Instructional	0.00 (0.00)	[0.00, 0.00]	0.00 (0.00)	[0.00, 0.00]	0.35 (1.34)	[-0.16, 0.88]
Frustrational	0.09 (0.30)	[-0.11, 0.29]	0.00 (0.00)	[0.00, 0.00]	0.00 (0.00)	[0.00, 0.00]
Two Levels Below						
Independent	0.00 (0.00)	[0.00, 0.00]	-0.08 (0.29)	[-0.27, 0.10]	0.96 (4.15)	[-0.64, 2.57]
Instructional	0.00 (0.00)	[0.00, 0.00]	0.00 (0.00)	[0.00, 0.00]	0.07 (0.26)	[-0.03, 0.17]
Frustrational	0.36 (0.81)	[-0.18, 0.91]	0.00 (0.00)	[0.00, 0.00]	0.04 (0.19)	[-0.04, 0.11]
Three Levels Below						
Independent	0.09 (0.30)	[-0.11, 0.29]	-0.17 (0.58)	[-.053, 0.20]	2.18 (11.33)	[-2.22, 6.57]
Instructional	0.27 (0.91)	[-0.33,0.88]	0.17 (0.39)	[-0.08, 0.41]	0.00 (0.00)	[0.00, 0.00]
Frustrational	1.36 (3.91)	[-1.26, 3.99]	0.00 (0.43)	[-0.27, 0.27]	-0.04 (0.19)	[-0.11, 0.04]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero). No *p*-values were < .05 from one-sample *t*-test analyses, indicating that student scoring was not significantly different than accurate scoring of number of words read. (No *p*-values were < .05 from Wilcoxon Signed Rank Test Analyses.) 0.00 (0) indicates *t*-test could not be run because standard deviation was zero (all students scored this passage correctly).

Table 4
Study 1: Student Identification of Reading Miscues

	8th Grade (*N* = 11)		7th Grade (*N* = 12)		6th Grade (*N* = 28)	
	M (SD)	95% CI	*M (SD)*	95% CI	*M (SD)*	95% CI
On Grade Level						
Independent	4.09 (1.51)*	[3.07, 5.11]	3.08 (1.00)*	[2.45, 3.72]	3.04 (1.29)*	[2.54, 3.54]
Instructional	7.09 (1.22)*	[6.27, 7.91]	4.00 (1.71)*	[2.92, 5.08]	5.11 (2.54)*	[4.12, 6.09]
Frustrational	5.00 (1.18)*	[4.21, 5.79]	4.67 (1.88)*	[3.48, 5.86]	3.00 (1.72)*	[2.33, 3.67]
One Level Below						
Independent	2.64 (1.43)*	[1.67, 3.60]	3.42 (1.38)*	[2.54, 4.29]	2.61 (1.34)*	[2.09, 3.13]
Instructional	4.18 (1.25)*	[3.34, 5.02]	5.50 (2.88)*	[3.67, 7.33]	4.18 (2.00)*	[3.40, 4.95]
Frustrational	3.09 (1.58)*	[2.03, 4.15]	2.67 (1.07)*	[1.98, 3.35]	2.64 (1.59)*	[2.03, 3.26]
Two Levels Below						
Independent	3.64 (1.12)*	[2.88, 4.39]	3.17 (1.64)*	[2.12, 4.21]	1.93 (1.25)*	[1.45, 2.41]
Instructional	6.00 (1.10)*	[5.26, 6.74]	4.08 (1.78)*	[2.95, 5.22]	3.79 (1.77)*	[3.10, 4.47]
Frustrational	4.18 (1.25)*	[3.34, 5.02]	2.42 (1.00)*	[1.78, 3.05]	1.75 (1.27)*	[1.26, 2.24]
Three Levels Below						
Independent	2.18 (0.98)*	[1.52, 2.84]	1.67 (1.44)*	[0.75, 2.58]	1.86 (1.65)*	[1.22, 2.50]
Instructional	4.73 (1.49)*	[3.73, 5.73]	4.00 (2.05)*	[2.70, 5.30]	1.82 (1.09)*	[1.40, 2.24]
Frustrational	3.09 (1.14)*	[2.33, 3.85]	1.67 (1.37)*	[0.80, 2.54]	0.50 (0.92)*	[0.14, 0.86]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero).
**p* values < .05 from one-sample *t*-test analyses, indicating that student identification of reading miscues was significantly different than accurate scoring. (All passages had *p*-values < .05 on Wilcoxon Signed Rank Tests.)

Table 5
Percentage of Passages With Accurate Counting or Computation

	Accurate Counting of Words Read (mistakes on 2 or fewer passages)	Accurate Counting of Marked Miscues (mistakes on 2 or fewer passages)	Accurate Computation of CWPM (mistakes on 2 or fewer passages)
Study 1			
Students			
6th Grade (*N* = 28)	89%	96%	89%
7th Grade (*N* = 12)	83%	100%	100%
8th Grade (*N* = 11)	91%	100%	82%
Study 2			
Students			
6th Grade (*N* = 21)	81%	90%	95%
7th Grade (*N* = 13)	69%	100%	69%
8th Grade (*N* = 16)	50%	75%	94%
Student Teachers			
Undergrad (*N* = 26)	46%	100%	100%
Graduate (*N* = 26)	100%	96%	100%

Note. CWPM: Correct Words Per Minute. Passages were counted as "accurate" if there were 2 or fewer errors in counting (words read and/or counting miscues) or computation of CWPM (i.e., subtraction) based on student notations on scoring protocols.

graders (*N* = 50). The majority of participants were male (58%); student ethnicities were Caucasian (78%), African American (13%), Hispanic (4%), and mixed ethnicities (4%). Table 2 summarizes achievement data for these students by grade level.

Student teachers. Fifty-two undergraduate (*n* = 26) and graduate (*n* = 26) students enrolled in two special education teacher preparation courses volunteered to participate in the study. A majority were female (94%) and Caucasian (98%). None of the undergraduate students had any years of teaching experience; graduate students' teaching experience ranged from 2 to 27 years. Only 1 of the 52 participants reported having learned about CBM in a previous course.

Training procedures. Two adjustments were made to the training materials in Study 2 based on our experience from the first experiment. First, the materials were revised to shorten the duration required to deliver the training and better focus the participants on the scoring conventions within a shorter amount of class time. Specifically, all participants (students and student teachers) received a more focused and streamlined 20-minute training session on the scoring conventions for ORF (described under General Procedures and streamlined based on the experiences in Study 1). However, to streamline the training, only the most frequently observed scoring conventions were included in the training: mispronounced words, omissions, and hesitations.

Second, a certification test was developed and used to determine proficiency in (a) notating scoring errors within passages, (b) counting words read and reading miscues, and (c) calculating CWPM. The certification test was added based on observed errors made in these areas by participants in Experiment 1. After all students had demonstrated proficiency in the individual errors in isolation, students were provided an opportunity to score a short paragraph containing 3-4 sentences. Finally, participants were required to demonstrate 100% accuracy on the certification test prior to moving forward; students who did not reach 100% were provided one-on-one reteaching until achieving 100% accuracy on the certification test.

Student teachers received the same training as the students but in an online format. This presentation format was selected because past research (Riccomini & Stecker, 2005) demonstrated its effectiveness for training preservice teachers to perform reading fluency assessments. Student teacher participants did not have to take the certification test

because it was assumed that adults would not have problems with counting or basic calculation. After training, all 52 student teachers rated the online training as either useful or very useful.

Results

Student scoring. Findings from analysis of student accuracy in scoring words read in one minute produced insufficient evidence to conclude that student scoring errors were significantly different from zero (i.e., accurate scoring). Further, statistical analysis could not be conducted on 15 of 36 passages (42%) because all students scored the passages accurately. Inspection of confidence intervals showed that upper and lower ranges were within acceptable ranges on all passages (see Table 6).

Statistically significant differences between student scoring of reading miscues and zero for each of the passages were found, except for one. However, inspection of confidence intervals showed that upper and lower ranges were only marginally outside acceptable ranges on all passages (see Table 7).

Student teacher scoring. Findings from analysis of student teacher accuracy in scoring words read in one minute produced insufficient evidence to conclude that errors made by graduate-level student teachers were significantly different from zero, and only one passage was significantly different for undergraduate student teachers. Further, statistical analysis could not be conducted on 12 of 18 passages (67%) for graduate student teachers and 13 of 18 passages (72%) for undergraduate student teachers, because all student teachers scored the passage accurately. Inspection of confidence intervals showed that upper and lower ranges were within acceptable ranges on all passages (see Table 8).

Statistically significant differences between student teacher scoring of reading miscues and zero (accurate scoring) were obtained for both 44% of passages scored by graduate student teachers and 78% of passages scored by undergraduate student teachers, indicating that graduate teachers were more accurate than undergraduate teachers. However, inspection of confidence intervals showed that upper and lower ranges were only marginally

Table 6
Study 2: Student Identification of Words Read

	8th Grade (*N* = 16)		7th Grade (*N* = 13)		6th Grade (*N* = 21)	
	M (SD)	95% CI	*M (SD)*	95% CI	*M (SD)*	95% CI
On Grade Level						
Independent	-0.69 (2.75)	[-2.15, 0.78]	0.62 (1.50)	[-0.29, 1.52]	1.33 (6.11)	[-1.45, 4.11]
Instructional	0.00 (0)	[0.00, 0.00]	-1.77 (4.90)	[-4.73, 1.19]	0.00 (0)	[0.00, 0.00]
Frustrational	-0.53 (2.07)	[-1.68, 0.61]	0.00 (0)	[0.00, 0.00]	0.14 (0.48)	[-0.07, 0.36]
One Level Below						
Independent	0.80 (1.82)	[-0.21, 1.81]	0.00 (0)	[0.00, 0.00]	-0.14 (0.66)	[-0.44, 0.16]
Instructional	0.44 (1.03)	[-0.11, 0.99]	0.00 (0)	[0.00, 0.00]	0.48 (1.63)	[-0.27, 1.22]
Frustrational	-0.50 (2.28)	[-1.72, 0.72]	-0.15 (0.56)	[-0.49, 0.18]	0.00 (0)	[0.00, 0.00]
Two Levels Below						
Independent	0.00 (0)	[0.00, 0.00]	-0.38 (1.39)	[-1.22, 0.45]	0.19 (0.75)	[-0.15, 0.53]
Instructional	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]	0.05 (0.22)	[-0.05, 0.15]
Frustrational	0.13 (0.50)	[-0.14, 0.39]	0.00 (0)	[0.00, 0.00]	0.05 (0.22)	[-0.05, 0.15]
Three Levels Below						
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	0.19 (0.75)	[-0.21, 0.59]	0.08 (0.28)	[-0.09, 0.24]	1.00 (2.65)	[-0.20, 2.20]
Frustrational	0.00 (0)	[0.00, 0.00]	-0.08 (0.28)	[-0.24, 0.09]	0.00 (0)	[0.00, 0.00]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero). No *p*-values were < .05 from one-sample *t*-test analyses, indicating that student scoring was not significantly different than accurate scoring of number of words read. (No *p*-values were < .05 from Wilcoxon Signed Rank Test Analyses except for the 6th-grade instructional passage, $p = .024$.) 0.00 (0) indicates *t*-test could not be run because standard deviation was zero (all students scored this passage correctly).

Table 7
Study 2: Student Identification of Reading Miscues

	8th Grade (*N* = 16)		7th Grade (*N* = 13)		6th Grade (*N* = 21)	
	M (SD)	95% CI	*M (SD))*	95% CI	*M (SD)*	95% CI
On Grade Level						
Independent	3.69 (1.62)*	[2.82, 4.55]	2.38 (2.10)*	[1.11, 3.66]	3.81 (1.44)*	[3.16, 4.46]
Instructional	6.31 (1.74)*	[5.39, 7.24]	3.54 (1.98)*	[2.34, 4.74]	5.90 (3.10)*	[4.50, 7.31]
Frustrational	4.00 (2.16)*	[2.85, 5.15]	2.85 (1.82)*	[1.75, 3.95]	3.29 (1.90)*	[2.42, 4.15]
One Level Below						
Independent	3.25 (1.34)*	[2.54, 3.96]	2.46 (1.71)*	[1.43, 3.50]	3.62 (1.20)*	[3.07, 4.17]
Instructional	3.75 (1.61)*	[2.89, 4.61]	4.85 (1.77)*	[3.78, 5.92]	4.90 (1.81)*	[4.08, 5.73]
Frustrational	3.06 (1.98)*	[2.01, 4.12]	1.92 (1.66)*	[0.92, 2.92]	2.62 (1.50)*	[1.94, 3.30]
Two Levels Below						
Independent	3.13 (1.50)*	[2.33, 3.92]	2.08 (1.71)*	[1.05, 3.11]	2.43 (1.29)*	[1.84, 3.01]
Instructional	4.94 (2.21)*	[3.76, 6.11]	2.69 (1.70)*	[1.66, 3.72]	5.33 (1.43)*	[4.68, 5.98]
Frustrational	2.50 (1.75)*	[1.57, 3.43]	2.46 (1.39)*	[1.62, 3.30]	2.10 (0.94)*	[1.67, 2.52]
Three Levels Below						
Independent	2.69 (1.49)*	[1.89, 3.48]	0.46 (2.40)	[-0.99, 1.91]	2.33 (1.56)*	[1.62, 3.04]
Instructional	3.69 (1.54)*	[2.87, 4.51]	2.46 (1.76)*	[1.40, 3.53]	2.33 (1.24)*	[1.77, 2.90]
Frustrational	2.63 (0.81)*	[2.20, 3.05]	1.62 (1.12)*	[0.94, 2.29]	0.76 (1.00)*	[0.31, 1.21]

Note. Numbers in table indicate deviation from correct scoring (i.e., zero).
*p values < .05 from one-sample t-test analyses, indicating that student identification of reading miscues was significantly different from accurate scoring. (All passages had p-values < .05 on Wilcoxon Signed Rank Tests.)

outside acceptable ranges on all passages for both groups (see Table 9).

Counting and computation accuracy. Counting and subtraction errors made by participants were tallied to determine accuracy (see Table 5). Overall, student teachers were more accurate than students. When counting the number of *words read*, students (in both studies) were not as accurate as the graduate program student teachers in Study 2, but they were more accurate that the undergraduate student teachers. Almost half of the undergraduate student teachers miscounted by one word (suggesting they did not understand how to use the preprinted numbers indicating cumulative numbers of words for each line). When counting the number of marked *reading miscues*, students in Study 1 were as accurate as both graduate and undergraduate student teachers, but sixth and eighth graders in Study 2 were not as accurate as the participating student teachers. When subtracting to compute CWPM, only seventh graders in Study 1 were as accurate as the student teachers.

General Discussion

In both studies, findings showed that after a focused training in ORF assessment scoring, students were highly accurate in identifying *words read* in one minute regardless of grade level (on grade or one, two, and three grades below) or instructional level (independent, instructional, frustration). This finding was the same for both graduate and undergraduate student teachers in Study 2. Because determining words read in an ORF assessment requires counting skills, we documented accuracy here and found that average student performance varied (Study 1 = 88%; Study 2 = 67%), as did student teacher performance (undergraduate teachers = 46%; graduate teachers = 100%).

In both studies, students were less consistent in correctly identifying *reading miscues* regardless of grade level (on grade, or one, two, and three grades below) or instructional level (independent, instructional, frustration). However, inspection of the findings showed that student teachers were not consistently accurate in scoring miscues either. Determining errors made in an ORF assessment also requires counting skills; students were highly accu-

rate in both studies (Study 1 = 96%; Study 2 = 88%) as were the student teachers (undergraduate teachers = 100%; graduate teachers = 96%). Further, determining CWPM in an ORF assessment requires basic computation. Findings here showed that, on average, students were highly accurate (Study 1 = 90%; Study 2 = 86%), as were the student teachers (graduate and undergraduate = 100%).

Implications for Practice

These combined findings provide initial support for the use of PAAR to conduct ORF measures as a screening for struggling readers at the middle school level, where this type of measure is typically not feasible when administered one-on-one by the classroom teacher. In addition to having the potential to serve as a screening tool to identify individual students who may be struggling, PAAR also has the potential to help teachers in selecting instructional materials and designing classwide instruction.

Although the use of PAAR in middle school may be more feasible than assessing individual students, it is not intended as a comprehensive measure to be used for educational decisions such as placement or formal evaluation. The overarching goal for the design and implementation of this research was to determine if peers could assess ORF with acceptable levels of reliability that would provide teachers with useful information for immediate classroom instructional

Table 8
Study 2: Student Teacher Identification of Words Read

	Graduate Level		Undergraduate Level	
	M (SD)	95% CI	*M (SD)*	95% CI
8th Grade				
Independent	0.00 (0)	[0.00, 0.00]	5.38 (15.20)	[-7.33, 18.08]
Instructional	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Frustrational	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
7th Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Frustrational	0.11 (0.32)	[-0.05, 0.27]	0.13 (0.34)	[-0.06, 0.31]
6th Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.04 (0.20)	[-0.04, 0.12]
Instructional	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Frustrational	0.00 (0.30)	[-0.12, 0.12]	0.00 (0)	[0.00, 0.00]
5th Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	0.23 (0.82)	[-0.10, 0.56]	0.12 (0.59)	[-0.12, 0.35]
Frustrational	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
4th Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	0.18 (0.39)	[-0.03, 0.38]	0.22 (0.43)*	[0.01, 0.43]
Frustrational	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
3rd Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	-0.13 (0.35)	[-0.42, 0.17]	0.00 (0)	[0.00, 0.00]
Frustrational	-0.13 (0.35)	[-0.42, 0.17]	0.00 (0)	[0.00, 0.00]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero). 0.00 (0) indicates *t*-test could not be run because standard deviation was zero (all teachers scored this passage correctly).
*p values < .05 from one-sample *t*-test analyses, indicating that teacher identification of words read was significantly different from accurate scoring. (Identical results were found in analysis using Wilcoxon Signed Rank Tests.)

Table 9
Study 2: Student Teacher Identification of Reading Miscues

	Graduate Level		Undergraduate Level	
	M (SD)	95% CI	*M (SD)*	95% CI
8th Grade				
Independent	0.78 (1.09)	[-0.06, 1.62]	1.88 (1.73)*	[0.43, 3.32]
Instructional	2.11 (0.78)*	[1.51, 2.71]	2.13 (1.25)*	[1.08, 3.17]
Frustrational	1.89 (0.60)*	[1.43, 2.35]	2.88 (0.64)*	[2.34, 3.41]
7th Grade				
Independent	0.11 (1.45)*	[0.83, 0.61]	0.06 (1.48)*	[-0.73, 0.85]
Instructional	0.39 (1.04)	[-0.34, 0.45]	0.31 (0.79)*	[0.11, 0.74]
Frustrational	0.33 (1.24)	[-0.28, 0.95]	0.75 (0.93)*	[0.25, 1.25]
6th Grade				
Independent	0.08 (0.56)	[-0.15, 0.30]	0.50 (0.91)*	[0.13, 0.87]
Instructional	0.00 (1.47)	[-0.59, 0.59]	1.54 (2.02)*	[0.72, 2.36]
Frustrational	0.46 (0.72)*	[0.15, 0.76]	0.29 (0.49)	[0.17, 0.74]
5th Grade				
Independent	0.23 (0.95)	[-0.15, 0.61]	0.85 (1.35)*	[0.30, 1.39]
Instructional	-0.23 (0.99)	[-0.63, 0.17]	0.42 (1.17)	[-0.05, 0.90]
Frustrational	0.46 (0.91)*	[0.10, 0.83]	0.81 (1.02)*	[0.40, 1.22]
4th Grade				
Independent	0.06 (0.43)	[-0.16, 0.28]	0.00 (0.59)	[-0.30, 0.30]
Instructional	1.41 (1.23)*	[0.78, 2.04]	1.83 (1.42)*	[1.12, 2.54]
Frustrational	1.53 (0.51)*	[1.26, 1.79]	1.72 (0.58)*	[1.44, 2.01]
3rd Grade				
Independent	1.88 (1.25)*	[0.83, 2.92]	2.56 (0.53)*	[2.15, 2.96]
Instructional	0.63 (1.41)	[-0.55, 1.80]	0.78 (0.67)*	[0.27, 1.29]
Frustrational	0.25 (0.46)	[-0.14, 0.64]	0.11 (0.33)	[-0.15, 0.37]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero).
*p values < .05 from one-sample t-test analyses, indicating that teacher identification of words read was significantly different from accurate scoring. (Identical results were found in analysis using Wilcoxon Signed Rank Tests.)

decisions (e.g., use of assistive technology, alternative reading sources, small-group instruction). We do not recommend using peers for formal assessments, although our results indicated student peers were more accurate than preservice teachers.

Teachers will have to provide training to students prior to conducting peer assessment, which may be time consuming. As middle school teachers consider the use of peers for assessing ORF, they should carefully consider and weigh the cost-to-benefit ratio; our results indicate the potential benefits of using peers as a screening measure in middle school grades.

Another important and surprising finding was that middle school students were more accurate than the preservice student teacher participants. This could be explained through anecdotal observations made of the preservice teachers. These participants were taking an entire course on teaching reading where significant time was allocated to assessing reading beyond correct and incorrect and recording the type of miscue (e.g., mispronunciations, omissions, and hesitations), consistent with scoring conventions of oral reading fluency (see Gillet et al., 2017).

It is possible that the preservice teachers were trying to record too much information in the early stages of practicing measuring ORF, which negatively impacted their accuracy. The middle school participants, on the other hand, were just recording correct or incorrect, and this may have resulted in the higher accuracy. Either way, the training neces-

sary to achieve high levels of accuracy in measuring ORF along with specific miscues used in this research project was likely not sufficiently intensive.

When screening to identify struggling readers, accuracy of *words read* in a minute has practical importance and may be more important than identification of specific miscues. For example, through peer-assisted assessment in a seventh-grade social studies class, Suzie may identify that Johnny reads 86 CWPM. Because the peer is more likely not to identify all miscues made by the reader, Johnny may actually read 82 CWPM (i.e., Suzie failed to identify 4 errors). However, considering that a typical seventh grader reads approximately 150 CWPM, the educational impact is the same – the student is reading far fewer CWPM with grade-level material than expected for his or her grade. In this example, regardless of the peer's failure to identify all of the reader's miscues, a classroom teacher would immediately see that this is a student who is potentially at risk and should be monitored more closely by a teacher. As with any screening measure, the result should be considered as a single data point that can be used to make decisions about a student (Hasbrouck & Tindal, 2006).

Clearly, data obtained from PAAR is not appropriate for formal identification or progress monitoring purposes if obtained from student scoring. However, these findings have implications for content-area classrooms at the middle school level. If utilized classwide, content-area teachers could potentially teach students how to score ORF assessments and obtain data on students in all of their classes in an extremely short period of time. This information could then be used both to inform instruction and to identify students whose progress should be closely monitored.

Conclusions

Nationally, reading performance continues to be a major concern. By the time students reach middle school, they are expected to read purposefully and comprehend, but many fail to do so (Ivey & Broaddus, 2000; Troia, 2006). In 2017, 37% of fourth-grade students performed at or above the *Proficient* level in reading skills, and only 36% of eighth-grade students performed at or above the *Proficient* level (National Center for Educational Statistics, 2018). Because today's U.S. classrooms are more diverse than ever before, students' reading levels vary widely in a single classroom, and the higher the grade level, the wider the range of reading abilities. To deliver quality instruction to meet the individual needs of all students in the classroom, teachers need to be aware of their students' current levels of performance.

Because middle school teachers often do not hear students read aloud, they may be unaware of which students in their classes are struggling with the assigned reading material. Fluency assessments are efficient, reliable, and valid indicators of reading proficiency when used as screening measures (Fuchs et al., 2001; Hasbrouck & Tindal, 2006). Knowing which students have the most difficulty in reading a textbook fluently can help teachers select appropriate texts for homework and instruction, match students to appropriate modifications (e.g., use of a text-to-speech version of the content), tailor instructional approaches to meet the needs of entire classes, and/or intensify instruction to address the reading capabilities of all students in the class (Berkeley & Lindstrom, 2011; Riccomini, Morano, & Hughes, 2017). Findings from this study suggest that using PAAR as a screening assessment has potential to help teachers make these important instructional decisions.

References

American Recovery and Reinvestment Act (ARRA) of 2009, Pub. L. No. 111-5, 123 Stat. 115, 516 (2009).

Barth, A. E., Stuebing, K. K., Fletcher, J. M., Denton, C. A., Vaughn, S., & Francis, D. (2014). The effect of reading duration on the reliability and validity of middle school students' ORF performance. *Assessment for Effective Intervention : official journal of the Council for Educational Diagnostic Services*, *40*(1), 53-64. https://doi.org/10.1177/1534508414545643

Berkeley, S., & Lindstrom, J. H. (2011). Technology for the struggling reader: Free and easily accessible resources. *Teaching Exceptional Children*, *43*, 48-55.

Berkeley, S., & Riccomini, P. J. (2017). Academic progress monitoring. In J. M. Kauffman & D. P. Hallahan (Eds.), *Handbook of special education* (2nd ed., pp. 218-232). New York, NY: Routledge. https://doi.org/10.1177/004005991104300405

Brigham, F. J., Berkeley, S., & Walker, E. (2012). Navigating the measurement and monitoring maze. *Exceptionality, 20,* 133-142. doi:10.1080/09362835.2012.669298

Brigham, F. J., Tochterman, S., & Brigham, M.S.P. (2000). Students with emotional and behavioral disorders and their teachers in test-linked systems of accountability. *Assessment for Effective Intervention, 26*(1), 19-27. https://doi.org/10.1177/073724770002600104

Calhoon, M. B., & Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial & Special Education, 24,* 235-245. https://doi.org/10.1177/07419325030240040601

Calkins, L., Ehrenworth, M., & Lehman, C. (2012). *Pathways to the common core: Accelerating achievement*. Portsmouth, NH: Heinemann.

Deno, S. L. (2003). Curriculum-based measures: Development and perspectives. *Assessment for Effective Intervention, 28*(3-4), 3-12. https://doi.org/10.1177/073724770302800302

Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., … Francis, D. J. (2011). The relations among oral reading and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading, 15,* 109-135. https://doi.org/10.1080/10888431003623546

Espin, C. A., & Deno, S. L. (1995). Curriculum-based measures for secondary students: Utility and task specificity of text-based reading and vocabulary measures for predicting performance on content-area tasks. *Diagnostique, 20,* 121-142.

Espin, C. A., Shin, J., & Busch, T. W. (2005). Curriculum-based measurement in the content areas: Vocabulary-matching as an indicator of social studies learning. *Journal of Learning Disabilities, 38,* 353-363. https://doi.org/10.1177/00222194050380041301

Espin, C. A., Wallace, T., Lembke, E., Campbell, H., & Long, J. D. (2010). Creating a progress-monitoring system in reading for middle-school students: Tracking progress toward meeting high-stakes standards. *Learning Disabilities Research & Practice,* 25, 60-75. https://doi.org/10.1111/j.1540-5826.2010.00304.x

Every Child Succeeds Act. (2015). Public Law No. 114-95. Available from https://www.congress.gov/bill/114th-congress/senate-bill/1177.

Fuchs, D., Fuchs, L., & Compton, D. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78,* 263-279. https://doi.org/10.1177/001440291207800301

Fuchs, L. S., Fuchs, D., Hosp, M., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239-256. https://doi.org/10.1207/S1532799XSSR0503_3

Fuchs, L., Fuchs, D., & Kazdan, S. (1999). Effects of peer-assisted learning strategies on high school students with serious reading problems. *Remedial & Special Education, 20,* 309-318. https://doi.org/10.1177/074193259902000507

Gillet, J. W., Temple, C. A., Temple, C. N., & Crawford, A. N. (2017). *Understanding reading problems* (9th ed.). Boston, MA: Pearson.

Greenwood, C. R., Maheady, L., & Delquadri, J. C. (2002). Class-wide peer tutoring. In G. Stoner, M. R. Shinn, & H. Walker (Eds.), *Interventions for achievement and behavior problems* (2nd ed., pp. 611-649). Washington, DC: National Association of School Psychologists.

Harniss, M. K., Caros, J., & Gersten, R. (2007). Impact of the design of U.S. history textbooks on content acquisition and academic engagement of special education students: An experimental investigation. *Journal of Learning Disabilities, 40,* 100-110. https://doi.org/10.1177/00222194070400020101

Harris, R., Marchand-Martella, N., & Martella, R. (2000). Effects of a peer-delivered corrective reading program. *Journal of Behavioral Education, 10,* 21-36. https://doi.org/10.1023/A:1016691705754

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59,* 636-644. https://doi.org/10.1598/RT.59.7.3

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42,* 963-974. doi:10.1037/0003-066x.42.11.963

Hiebert, E. H. (2009). The (mis)match between texts and students who depend on schools to become literate. In E. H. Hiebert & M. Sailors (Eds.), *Finding the right texts: What works for beginning and struggling readers* (pp. 1-20). New York, NY: Guilford Press.

Hosp, M. K., & Hosp, J. L. (2003). Curriculm-based measurement for reading, spelling, and math: How to do it and why. *Preventing School Failure, 48,* 10-17. https://doi.org/10.1080/1045988X.2003.10871074

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review, 34*(1), 9-26.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement* (2nd ed.). New York, NY: Guilford Press.

Individuals With Disabilities Improvement Act of 2004, 20 U.S.C. § 1400 *et seq.* (2005).

Ivey, G., & Broaddus, K. (2000). Tailoring the fit: Reading instruction and middle school readers. *The Reading Teacher, 54,* 68-78.

Jenkins, J. R. (2009). Measuring reading growth: New findings on progress monitoring (PM). *New Times for DLD, 27*(1), 1-2.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formatove teaching: Reading Aloud and Maze. *Exceptional Chirldren, 59*(5), 421-432. https://doi.org/10.1177/001440299305900505

Kendall, J., Ryan, S., Alpert, A., Richardson, A., & Schwols, A. (2012). *State adoption of the comon core state standards: The 15 percent rule.* Denver, CO: Mid-Continent Research for Education and Learning.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95,* 3-21. https://doi.org/10.1037/0022-0663.95.1.3

Lembke, E. S., Hampton, D., & Hendricker, E. (2013). Data-based decision making in academics using curriculum-based measurment. In J. W. Lloyd, T. J. Landrum, B. G. Cook, & M. Tankersley (Eds.), *Research-based approaches for assessment* (pp. 18-31). Boston, MA: Pearson.

MacGinitie, W. H., MacGinitie, R., Maria, R., & Dreyer, L. (2000). *Gates-MacGinitie Reading Tests.* Rolling Meadows, IL: Riverside Publishing.

National Center for Educational Statistics. (2018). *NAEP reading report card.* Retrieved from https://www.nationsreportcard.gov/.

Maheady, L., Harper, G. F., & Mallette, B. (2003). A focus on class wide peer tutoring. *Current Practice Alerts, 8.*

Marshak, L., Mastropieri, M. A., & Scruggs, T. E. (2011). Curriculum enhancements in inclusive secondary social studies classrooms. *Exceptionality, 19*, 61-74. https://doi.org/10.1080/09362835.2011.562092

Mastropieri, M. A., Scruggs, Mohler, L., Beranek, M., Boon, R., Spencer, V., & Talbott, E. (2001). Can middle school students with serious reading difficulties help each other and learn anything? *Learning Disabilities Research & Practice, 16*, 18-27. https://doi.org/10.1111/0938-8982.00003

Meisinger, E. B., Bloom, J. S., & Hynd, G. W. (2010). Reading fluency: Implications for the assessment of children with reading disabilities. *Annals of Dyslexia, 60(1)*, 1-17. doi:10.1007/s11881-009-0031-z

Mercer, C. D., Campbell, K. U., Miller, M., Mercer, K. D., & Lane, H. B. (2000). Effects of a reading fluency intervention for middle schoolers with specific learning disabilities. *Learning Disabilities Research & Practice, 15,* 179-189. https://doi.org/10.1207/SLDRP1504_2

Morano, S., & Riccomini, P. J. (2017). Re-examining the literature: The impact of peer tutoring on higher order learning. *Preventing School Failure: Alternative Education for Children and Youth, 61*(2), 104-115. doi: 10.1080/1045988X.2016.1204593

Neddenriep, C. E., Fritz, A. M., & Carrier, M. E. (2011). Assessing for generalized improvements in reading comprehension by intervening to improve reading fluency. *Psychology in the Schools, 48*(1), 14-27. doi:10.1002/pits.20542

Petscher, Y., Cummings, K. D., Biancarosa, G., & Fien, H. (2013). Advanced (measurement) applications of curriculum-based measurement in reading. *Assessment for Effective Intervention, 38*(2), 71-75. doi:10.1177/1534508412461434

Read Naturally. (2008a). *Reading fluency monitor*. Saint Paul, MN: Read Naturally

Read Naturally. (2008b). *Reading fluency benchmark assessor*. Saint Paul, MN: Read Naturally.

Riccomini, P. J., Morano, S., & Hughes, C. A. (2017). Big ideas in special education: Specially designed instruction, high-leverage practices, explicit instruction, and intensive instruction. *Teaching Exceptional Children, 50*(6), 1-8. https://doi.org/10.1177/0040059917724412

Riccomini, P. J., & Stecker, P. M. (2005). Effects of technology-enhanced practice on scoring accuracy of oral reading fluency. *Journal of Special Education Technology, 20*(3), 5-12. https://doi.org/10.1177/016264340502000301

Roach, A. T. (2008). Intervention validity of assessments used in special education: Introduction to the special series. *Assessment for Effective Intervention, 34*(1), 3-5. doi:10.1177/1534508408314109

Santi, K., & Vaughn, S. (2007). Progress monitoring: an integral part of instruction. *Reading & Writing, 20*, 535-537. https://doi.org/10.1007/s11145-007-9053-2

Schwanenflugel, P. J., & Ruston, H. P. (2008). Becoming a fluent reader: From theory to practice. In M. R. Kuhn & P. J. Schwanenflugel (Eds.), *Fluency in the classroom* (pp. 1-16). New York, NY: Guilford Press.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationships of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43*(5), 527-535. doi:10.1002/pits.20175

Siskin, L. (1994). *Realms of knowledge: Academic departments in secondary schools*. Bristol, PA: Falmer Press.

Soriano, M., Miranda, A., Soriano, E., Nievas, F., & Felix, V. (2011). Examining the efficacy of an intervention to improve fluency and reading comprehension in Spanish children with reading disabilities. *International Journal of Disability, Development and Education, 58*(1), 47-59. doi:10.1080/1034912x.2011.547349

Spencer, V. G., Scruggs, T. E., & Mastropieri, M. A. (2003). Content area learning in middle school social studies classrooms and students with emotional or behavioral disorders: A comparison of studies. *Behavioral Disorders, 28*, 77-93. https://doi.org/10.1177/019874290302800208

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. https://doi.org/10.1002/pits.20113

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (2012). *TOWRE: Test of Word Reading Efficiency.* Austin, TX: Pro-Ed.

Troia, G. A. (2006). Meaningful assessment of content-area literacy for youth with and without disabilities. *Assessment for Effective Intervention, 31*, 69-80. https://doi.org/10.1177/073724770603100206

Venn, J. (2013). Curriculum-based assessment. In J. W. Lloyd, T. J. Landrum, B. G. Cook, & M. Tankersley (Eds.), *Research-based approaches for assessment* (pp. 84-93). Boston, MA: Pearson.

Wayman, M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41,* 85-120. https://doi.org/10.1177/00224669070410020401