Published in final edited form as:

Assess Eff Interv. 2014 December ; 40(1): 53–64. doi:10.1177/1534508414545643.

The Effect of Reading Duration on the Reliability and Validity of Middle School Students' ORF Performance

Amy E. Barth, PhD¹, Karla K. Stuebing, PhD², Jack M. Fletcher, PhD², Carolyn A. Denton, PhD³, Sharon Vaughn, PhD⁴, and David Francis, PhD²

¹University of Missouri, Columbia, USA

²University of Houston, TX, USA

³University of Texas Health Science Center at Houston, TX, USA

⁴University of Texas at Austin, USA

Abstract

We evaluated the technical adequacy of oral reading fluency (ORF) probes in which 1,472 middle school students with and without reading difficulties read fluency probes for 60 s versus reading the full passage. Results suggested that the reliability of 60-s probes ($r_s = .75$) was not substantively different than full passage probes ($r_s = .77$) among struggling readers and typically developing readers in Grades 6 to 8. The correlation of 60-s and the full passage probes with norm-referenced measures of ORF ranged from .32 to .83, and the correlation with norm-referenced measures of reading comprehension ranged from .32 to .54, indicating that both measures were moderately valid and adequate for use among middle school students. Last, full passage probes with sensitivity rates ranging from .40 to .45 were only slightly more sensitive for identifying at-risk readers than 60-s probes, with sensitivity rates ranging from .36 to .40, suggesting that the full passage probes identified a slightly higher percentage of at-risk students with reading difficulties.

Keywords

oral reading fluency; middle-grade readers; reliability; validity

© Hammill Institute on Disabilities 2014

Corresponding Author: Amy E. Barth, Department of Special Education, University of Missouri, 311B Townsend Hall, Columbia, MO 65211, USA. bartha@missouri.edu.

Reprints and permissions: sagepub.com/journalsPermissions.nav

Authors' Note

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Child Health and Human Development (NICHD), the National Institutes of Health, or the Institute of Education Sciences, U.S. Department of Education.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Oral Reading Fluency (ORF) in the Middle Grades

Recent research suggests that approximately 46% to 88% of struggling middle-grade readers (Grades 6–9) present significant ORF deficits (Brasseur-Hock, Hock, Kieffer, Biancarosa, & Deshler, 2011; Cirino et al., 2013; Hock et al., 2009). As a consequence, ongoing assessment of ORF is increasingly used to identify students with word reading accuracy/fluency difficulties and monitor reading progress in the middle grades (Crawford, Tindal, & Stieber, 2001; Hintze & Silbergitt, 2005; McGlinchey & Hixson, 2004; Nolet & McLaughlin, 2000; Silbergitt & Hintze, 2005; Wallace, Espin, McMaster, Deno, & Foegen, 2007). However, limited data on the reliability and validity of ORF assessment at the middle grades are available raising concerns about its technical adequacy (Wallace et al., 2007).

Technical Adequacy of ORF in the Middle Grades

Research on the technical adequacy of ORF in the middle grades indicates that the reliability and validity coefficients are moderate to high. Reliability coefficients for ORF, measured by calculating the number of words read correctly in 1 min (ORF-60), range from .87 to .92. Validity coefficients range from .42 to .62 (Espin & Foegen, 1996; Silbergitt, Burns, Madyun, & Lail, 2006; Torgesen, Nettles, Howard, & Winterbottom, 2005; Yovanoff, Duesbery, Alonzo, & Tindal, 2005).

Technical Adequacy of ORF Measured for Longer Durations Versus ORF-60

Although past research on the technical adequacy of ORF-60 supports its use with middle-grade readers, Espin, Wallace, Lembke, Campbell, and Long (2010) recently suggested that reading for longer durations (e.g., >1 min) might improve the reliability and validity of ORF assessments for older readers. To examine this, Daane, Campbell, Grigg, Goodman, and Oranje (2005) compared two methods for assessing reading rates among 1,779 students in Grade 4: (a) words per minute (WPM) for the first minute of reading and (b) average WPM for the full passage. Results indicated that the magnitude of the correlation between ORF scores calculated as the average WPM for the full passage and reading comprehension (validity) was higher than ORF scores calculated for the first minute. More recently, Ticha, Espin, and Wayman (2009) examined the reliability and validity of ORF calculated for three durations (1, 2, and 3 min of reading) among 35 students in Grade 8. Results revealed that ORF scores calculated across 2 or 3 min of reading were as reliable and valid as ORF scores calculated for 1 min of reading. Similarly, Espin et al. compared the technical adequacy of ORF calculated for three durations (1, 2, and 3 min of reading) among 238 students in Grade 8. Results revealed that reliability and validity was similar regardless of duration.

Limitations of Past Research

Both Ticha et al. (2009) and Espin et al. (2010) reported that among middle-grade readers, the reliability and validity of ORF measured for longer durations was not substantively different than the reliability and validity of ORF-60. In addition, Espin et al. reported that reading rates were invariant across different durations of reading. These findings contrast with Daane et al. (2005), who reported that ORF calculated for the full passage was associated with higher reading comprehension scores compared with ORF calculated for the

first minute. It may be that the differences in findings were due to differences in the technical adequacy of the ORF measures used across the three studies or that the impact of duration varies with grade level and potentially by reading ability, with younger students affected to a greater degree than older students and struggling readers affected to a greater degree than skilled readers.

Second, generalizations from these studies are limited due to the use of readability formulae to determine passage difficulty. Readability formulae imprecisely estimate the difficulty level of passages and are not adequate for equating passages (Ardoyn & Christ, 2009; Francis, Santi, et al., 2008). In the event that equating processes are not used, ORF scores are significantly influenced by contextual variables (i.e., administration order, text type, and text difficulty) and do not accurately reflect the true ORF abilities of the reader (Francis, Santi, et al., 2008).

Third, each study included students with reading disabilities and difficulties (in small proportions) but did not report reliability and validity coefficients separately for middle-grade struggling readers (see Espin & Deno, 1993a, 1993b; Espin & Foegen, 1996). Because ORF is increasingly used to monitor reading progress among students who perform at or below basic levels of reading on state reading accountability tests, a critical gap in research and practice is the technical adequacy of ORF among middle-grade struggling readers.

Research Questions

Given these concerns, the present study addressed two research questions that target the technical adequacy of ORF-60 and ORF for longer durations among middle-grade readers.

Research Question 1: Is the alternate form reliability and concurrent validity of ORF calculated for the first 60 s substantively different than oral reading fluency calculated for the full passage (ORF-FP) among struggling readers and typically developing readers in Grades 6 to 8?

We hypothesized that among middle-grade readers, the influence of reading duration on ORF rates varies by reader skill. We hypothesized that among struggling middle-grade readers, variance in ORF-60 scores will continue to be observed because they have not attained proficient levels of word reading accuracy and consequently may not have decoding skills adequate for the task. Therefore, the most reliable and valid index of ORF will be measured within the first minute. In contrast, we hypothesized that among skilled readers who are actively engaging language comprehension processes to integrate information in text and information in text with background knowledge; ORF measured for the full passage will more accurately quantify their ability to fluently read for meaning and will be reflected in higher reliability and validity coefficients.

The second research question involved classification accuracy, which is as follows:

Research Question 2: Is the predictive validity or classification accuracy (i.e., sensitivity, specificity, positive predictive rate, negative predictive rate, and overall classification rate) of ORF calculated for the first 60 s substantively different than ORF calculated for the full passage?

We hypothesized that the classification accuracy of ORF will exceed .85, thus providing middle school teachers with a practical, accurate, and efficient means of differentiating at-risk and not-at-risk readers with risk status aligning with proficient/not-proficient designations from state accountability reading assessments. This would also provide middle schools with a means of identifying risk status among new enrollees who lack scores on the state accountability reading assessment.

Method

Participants

School sites—This study was conducted in two large-urban communities in the southwestern United States, with approximately half the sample coming from each community. Enrollment at each middle school ranged from 633 to 1,300 students. Based on the state accountability rating system, two schools were rated as recognized, four as acceptable, and one school as academically unacceptable. Students qualifying for reduced or free lunch ranged from 56% to 86% in the first community and from 40% to 85% in the second community.

Students—The sample represented 1,472 sixth- to eighth-grade students from the seven schools during the 2006–2007 academic year. Students were first-year participants in a large multi-year study (see Vaughn, Cirino, et al., 2010; Vaughn, Wanzek, et al., 2010). The sample comprises 839 struggling readers and 633 typically developing readers. Struggling readers were defined as students who either (a) failed the reading comprehension component of the state accountability test (*Texas Assessment of Knowledge and Skills* [TAKS]; Texas Educational Agency [TEA], 2004), or (b) performed within one half of one standard error of measurement above the pass–fail cut-point on their first attempt in the spring of 2006 (i.e., scale scores ranging from 2100 to 2150 points). In addition, students in special education who did not take the reading subtest of TAKS but took the reading comprehension component of the *State Developed Alternative Assessment–II* (SDAA-II; TEA, 2004) were also defined as struggling readers (see Vaughn, Cirino, et al., 2010, for greater detail on the sample selection).

Typically developing readers were defined as students who scored greater than one half of one standard error of measurement above the pass–fail cut-point on the reading subtest of TAKS (i.e., scale scores above 2,150 points) on their first attempt in the spring of 2006. Students were excluded from the study if (a) they were enrolled in a special education life skills class; (b) they took the reading sub-test of the SDAA-II at a level lower than Grade 3; (c) they presented a significant visual, hearing, or intellectual disability); or (d) were classified as English as second language by their middle school and received primary instruction in a bilingual classroom. Students meeting these criteria were excluded because they would be unable to complete the assessments in the test battery. Because a large proportion of students passing TAKS (>80%), we randomly selected adequate readers within school (and grade) in proportion to the number of struggling readers. This resulted in a sample comprising 41% typically developing readers and 59% struggling readers.

Procedures

Testing procedures—Testing was completed at the student's middle school in quiet locations designated by the school (i.e., library, unused classrooms, theater, etc.). Students were assessed by examiners who completed an extensive training program conducted by the investigators regarding test administration and scoring procedures and met 95% accuracy on a fidelity checklist for each task in the assessment battery (e.g., accurate manipulation of testing materials, administration directions and item prompts, scoring, and packet verification, etc.). All data were verified for accuracy following testing and prior to data processing.

Testing timeline—Students were assessed at the beginning of the 2006 school year (end of September through October) with a battery that assessed reading fluency and reading comprehension abilities. The testing window spanned from September through October due to the large sample size. The testing window per students was a few days (range = 1–5 days). This was required given that some assessments were administered individually and others in a group, with both requiring scheduling during electives.

Measures

Rationale—To determine the reliability and validity of the ORF-60 and ORF-FP measures, outcome domains representing the constructs of ORF and reading comprehension were assessed. ORF measures included tests of sentence reading fluency, passage reading fluency, and timed decoding of real words and pseudowords. Reading comprehension was assessed with a traditional measure of reading comprehension in which students read several paragraphs and answered multiple choice questions that tapped both literal and inferential comprehension of text and a cloze procedure that assessed sentence-level comprehension.

ORF measures

Passage fluency (PF)—The PF (Francis, Barth, Cirino, Reed, & Fletcher, 2008) consists of 100 graded, expository ($n = 50$) and narrative ($n = 50$) passages, for use in Grades 6 to 8. Passages ranged from 108 words to 591 words in length (approximately 70% of passages were greater than 300 words), and ranged in difficulty from 350 to 1,400 Lexiles (Lexile Framework, 2007). The passages were derived from former TAKS, SDAA, and Texas Primary Reading Inventory passages and modified for length. Where there were gaps in grade level, passages were written by Language Arts teachers with extensive training on narrative and expository text structure. Written passages were reviewed by project researchers to ensure that content was appropriate for the targeted grade level using the TEA, Texas Essential Knowledge and Skills (TEKS) for English Language Arts and Reading for the Middle School. Passage Lexile was used to determine that the readability level or difficulty level was grade appropriate. All passages were then scaled to better equalize different forms specifically for this assessment of ORF using equipercentile equating. Equipercentile equating transformed each score on the ORF to a comparable score on a reference form (*Test of Word Reading Efficiency* [TOWRE] Sight Word Efficiency) that had the same percentile rank within the group of test-takers. Equating was carried out within grade and testing time points such that differences in mean ORF performance over

time and between grades were preserved. Past research using linearly equating method, reduced the effects of text type (i.e., expository and narrative), administration order, and text difficulty in the estimation of students' reading fluency abilities (Francis, Santi, et al., 2008). With equating, differences could not be attributed to older students reading less difficult passages, or students reading more difficult passages followed by reading easier passages later in the school year. For purposes of this study, students were administered five passages. For each passage, two ORF scores were generated: (a) reading rate representing the number of linearly equated words correct per minute (WCPM) for the first minute of reading (ORF-60) and (b) reading rate representing the average number of linearly equated words read correct per minute for the full passage (ORF-FP). Average ORF-60 and ORF-FP scores were calculated for the five passages administered.

Test of Silent Reading Efficiency and Comprehension (TOS-REC)—The TOSREC (Wagner, Torgesen, Rashotte, & Pearson, 2010) is a 3-min, group-based assessment of reading fluency and comprehension. Students were presented with a series of short sentences and were required to assess their veridicality. The raw score is the number of sentences correctly identified as true or false, minus the number of incorrect responses, within the 3-min time limit. The average internal consistency for middle school students in Grade 6 was .79 (Vaughn, Cirino, et al., 2010; Vaughn, Wanzek, et al., 2010). The standard score was the dependent measure used.

AIMSweb Maze Curriculum Based Measure (CBM) Reading Comprehension—The Maze CBM Reading Comprehension (Shinn & Shinn, 2002) subtest is a 3-min, group-based curriculum-based assessment of fluency and comprehension (Cirino et al., 2013). Students were presented with a 150 to 400 word passage. The first sentence is intact, but every seventh word of the remaining sentences is deleted. Students are required to identify the correct target word from among three choices while silently reading the text. The raw score is the number of targets correctly identified within the time limit and was the dependent measure used. Vaughn et al. (Vaughn, Cirino, et al., 2010; Vaughn, Wanzek et al., 2010) reported a mean intercorrelation of .79 in the Grade 6 sample of 327 struggling readers and 249 typical readers and a mean intercorrelation of .95 in the Grades 7 to 8 sample of 436 struggling readers and 440 typical readers.

TOWRE—Both Sight Word Efficiency and Phonemic Decoding Efficiency subtests (Form A) were administered at pretest (Torgesen, Wagner, & Rashotte, 1998). For the Sight Word Efficiency subtest, the student was given a list of 104 words and asked to read them as accurately and as quickly as possible. The raw score is the number of words read correctly within 45 s. For the Phonemic Decoding Efficiency subtest, the student was given a list of 63 non-words and asked to read them as accurately and as quickly as possible. The raw score is the number of nonwords read correctly within 45 s. Alternate form and test-retest reliability coefficients at or above .90 in this age range (Torgesen et al., 1998). The combined standard score for the two sub-tests was used.

Passage comprehension measures

Group Reading Assessment and Diagnostic Evaluation (GRADE)—The GRADE (Williams, 2001) is a group-based, norm-referenced, untimed test of reading comprehension. For *Passage Comprehension*, the students read several paragraphs and answer multiple choice questions that tap the skills of questioning, predicting, clarifying, and summarizing text. Among students in Grades 6 to 8, the coefficient alpha for the Passage Comprehension subtest ranged from .85 to .93 (fall) and .88 to .90 (spring) (Williams, 2001). The GRADE only provides a stanine standard score for the Passage Comprehension subtest. Rather than using this 9-point stanine-scaled score, the raw score for the Passage Comprehension subtest was used to derive a standard score for the GRADE Comprehension Composite, which is typically based on the Passage Comprehension and Sentence Comprehension measures. The prorated standard score represented the proportion of items correct/items administered. This estimated standard score was used in statistical analyses.

Woodcock Johnson–III Test of Achievement (WJ-III)—The WJ-III (Woodcock, McGrew, & Mather, 2001) is a nationally standardized individually administered battery of achievement tests (McGrew & Woodcock, 2001). Passage Comprehension uses a cloze procedure to assess sentence-level comprehension by requiring the student to read a sentence or short passage and fill in missing words based on the overall context. The coefficient alphas for students in Grades 6 to 8 exceed .90.

Reader subgroup measure

TAKS—The TAKS (TEA, 2004) is an untimed criteria-referenced reading comprehension test developed by Pearson Educational Measurement in conjunction with the TEA and is the Texas reading accountability test that was administered in February of each academic year. Different assessments are used for each grade, with each aligned to grade-based standards from the TEKS. Students read passages (both expository and narrative) and answer questions. After reading each passage, which typically has a title and illustrative pictures, students answer several multiple choice questions designed to access the literal meaning of the passage, vocabulary, and different aspects of critical reasoning about the material read. The internal consistency (coefficient alpha) of the Grade 7 test is .89 (TEA, 2004).

Analytic Approach

To address the alternate form reliability of ORF calculated for the first 60 s versus the full passage, we compared the correlations among passages 1 to 5 at pretest for ORF-60 and ORF-FP. To examine the concurrent validity of ORF-60 and ORF-FP scores, we computed correlations between average ORF-60 and ORF-FP scores at pretest and external measures of reading fluency (i.e., TOWRE, TOSREC, and AIMSweb Maze) and reading comprehension (i.e., GRADE Passage Comprehension and WJ-III Passage Comprehension) administered at the beginning of the school year. Reliability and validity statistics were reported for each grade and struggling versus typically developing readers within grade.

To address the second research question involving the predictive validity and classification accuracy of ORF for the first 60 s versus the full passage, we calculated sensitivity, specificity, positive predictive rate, negative predictive rate, overall classification accuracy,

and area under the Receiver Operating Characteristic (ROC) curve (using SAS 9.0 statistical software) for ORF-60 and ORF-FP. The fluency benchmark was the 25th percentile. This permitted comparisons of which score more accurately classified readers as at-risk or not-at-risk on the TAKS Reading Assessment. The 25th percentile represents ORF rates of 98, 102, and 106 WCPM in Grades 6, 7, and 8, respectively (Hasbrouck & Tindal, 2006). Risk status on the TAKS Reading Assessment followed the definition of struggling versus adequate reading using the cut-point of a standard score of 2,150.

Sensitivity, the extent to which a measure correctly classifies students at risk of reading difficulties as struggling readers, was calculated by dividing the number of true positives (those correctly identified as struggling readers) by the sum of true positives and false negatives (i.e., at-risk readers incorrectly classified as skilled readers). A sensitivity of 100% means that the test identifies all at-risk readers as struggling readers. Specificity, the extent to which a measure correctly classifies not-at-risk readers, was calculated by dividing the number of true negatives (students correctly identified as not-at-risk) by the sum of true negatives and false positives (skilled readers incorrectly classified as struggling readers). A specificity of 100% means that the test correctly classifies all skilled readers as proficient. The positive predictive rate was the proportion of students with a positive test result who were correctly identified as struggling readers, whereas the negative predictive rate was the proportion of students with a negative result who were correctly identified as skilled readers. Overall classification accuracy represented the proportion of correct classifications and was calculated by dividing the total number of correct classifications (true positives + true negatives) by the sum of correct and incorrect classifications (true positives + true negatives + false positives + false negatives). Area under the ROC curve (AUC) quantified the extent to which a measure correctly classified skilled and struggling readers, with an AUC of .5 representing chance classification and an AUC of 1.0 representing perfect classification.

Results

Preliminary Analyses

Prior to analyses, we evaluated distributional data, both statistically and graphically, for skewness, kurtosis, and normality, with few difficulties noted. None of the reading fluency variables exhibited significant skew or kurtosis. The WJ-III Passage Comprehension and GRADE Passage Comprehension were slightly kurtotic (i.e., 1.1 and 1.2, respectively). No students were missing data on the PF, AIMSweb Maze CBM Reading Comprehension, TOSREC, TOWRE, WJ-III Passage Comprehension, or TAKS. Twelve students were missing GRADE Passage Comprehension scores. All variables had a small number of outliers ($>3 SD$ from the mean). Analyses were performed with and without outliers, as well as with and without imputed scores for the 12 students missing GRADE Passage Comprehension scores. Because there were no differences in the pattern of results, the original data are reported in the analyses.

Demographic Comparisons and Descriptive Data

Table 1 summarizes demographic variables by grade and group. Chi-square analyses and ANOVAs were conducted to determine whether groups differed on demographic variables

within grade. Within Grades 6 to 8, struggling readers differed significantly from typically developing readers in subsidized lunch, race/ethnicity, and English as a second language (ESL) status ($ps < .001$), with a higher proportion of struggling readers receiving subsidized lunch, designated as ESL, and generally comprising racial/ethnic groups other than White compared with typically developing readers. However, these differences are small and the study is highly powered. There was only one significant interaction of reader group and grade on age, $F(5, 1466) = 613.0, p < .001$. Based on Tukey post hoc analyses ($p < .05$), struggling readers were older than typically developing readers within each grade, with differences between struggling readers and typically developing readers declining over Grades 6 to 8. This difference is most likely due to more frequent retention of struggling readers.

Table 2 reports the means and standard deviations for ORF-60 and ORF-FP as well as external measures of ORF and reading comprehension by reader subgroup within grade. Table 2 indicates that typically developing middle-grade readers performed at approximately the mean on standardized measures of reading comprehension and reading fluency; speeded decoding skills (i.e., TOWRE) were one-third standard deviation above the mean. Struggling readers performed approximately 1 standard deviation below the mean on measures of reading comprehension and also the TOSREC; speeded decoding skills were approximately one-half standard deviation below the mean.

Reliability

Typically developing readers—Among typically developing readers in Grade 6, the alternate form reliability of ORF-60 ranged from .82 to .89 (average $r = .86$); in Grade 7 from .78 to .89 (average $r = .85$); and in Grade 8 from .75 to .87 (average $r = .82$). For ORF-FP, the alternate form reliability among typically developing readers ranged from .84 to .91 (average $r = .87$) in Grade 6, .79 to .93 (average $r = .87$) in Grade 7, and from .85 to .91 (average $r = .88$) in Grade 8. Collapsing across Grades 6 to 8, alternate form reliabilities among the five ORF-60 scores ranged from .78 to .90, with alternate form reliabilities ranging from .79 to .93 for ORF-FP among typically developing readers.

Struggling readers—Among struggling readers in Grade 6, alternate form reliability of ORF-60 ranged from .84 to .88 (average $r = .86$); in Grade 7 from .84 to .90 (average $r = .87$); in Grade 8 from .88 to .92 (average $r = .89$). Alternate form reliability of ORF-FP in Grade 6 ranged from .77 to .90 (average $r = .85$), in Grade 7 from .80 to .92 (average $r = .87$), and in Grade 8 from .81 to .94 (average $r = .90$). Collapsing across grades, the alternate form reliabilities among the five ORF-60 scores ranged from .84 to .92 and .77 to .94 for ORF-FP for struggling readers. These reliability coefficients are similar for struggling and typical readers.

Concurrent Validity

Typically developing readers—Among typically developing readers, ORF-60 and ORF-FP correlated moderately well with external measures of reading fluency and reading comprehension across Grades 6 to 8. Correlations among ORF-60 and external measures of reading fluency ranged from .58 to .72 among typically developing students in Grade 6, .56

to .77 in Grade 7, and .48 to .62 in Grade 8. The correlations among ORF-60 GRADE Passage Comprehension ranged from .46 to .54 and WJ-III Passage Comprehension ranged from .43 to .49 across Grades 6 to 8. Correlations among ORF-FP and external measures of reading fluency ranged from .58 to .73 among typically developing readers in Grade 6, .56 to .77 in Grade 7, and .51 to .62 in Grade 8. The correlations among ORF-FP and GRADE Passage Comprehension ranged from .48 to .54 and WJ-III Passage Comprehension ranged from .43 to .49 across Grades 6 to 8.

Struggling readers—In Grade 6, the magnitude of the correlations among ORF-60 and external measures of reading fluency for struggling readers ranged from .39 to .79, with correlations among ORF-FP and external measures of reading fluency ranging from .39 to .79. In Grade 7, the magnitude of correlations among ORF-60 and external measures of reading fluency ranged from .32 to .83, with correlations among ORF-FP and measures of reading fluency ranging from .34 to .83. In Grade 8, the magnitude of the correlations among ORF-60 and ORF-FP and external measures of reading fluency ranged from .43 to .81. Across Grades 6 to 8, the magnitude of the relations among ORF-60 and ORF-FP and AIMswb Maze CBM Reading Comprehension tended to be lower among struggling readers than typically developing readers, whereas the relations among ORF-60 and ORF-FP and TOWRE tended to be higher among struggling readers than typically developing readers. Correlations among ORF-60 and Grade Passage Comprehension ranged from .32 to .45 and WJ-III Passage Comprehension ranged from .46 to .50 across Grades 6 to 8. Relations among ORF-FP and GRADE Passage Comprehension ranged from .34 to .49 and WJ-III Passage Comprehension ranged from .46 to .50 across Grades 6 to 8.

Classification Accuracy

Table 3 reports the sensitivity, specificity, positive predictive rate, negative predictive rate, and overall classification accuracy for ORF-60 and ORF-FP for the 25th percentile to determine whether ORF-60 is more or less sensitive at identifying at-risk middle-grade readers on the TAKS Reading Assessment.

Table 3 shows that ORF-60 correctly identified 40% (i.e., sensitivity) of students in Grade 6 with a reading difficulty and 92% (i.e., specificity) of those who do not have a reading difficulty based on the TAKS cut-point of 2,150. The false positive rate was 12% and false negative rate was 49%. Among students in Grades 7, ORF-60 correctly identified 38% of students with a reading difficulty and 94% of those who do not, with a false positive rate of 12% and a false negative rate of 44%. Among students in Grade 8, the 25th percentile on the ORF-60 correctly identified 36% of students with a reading difficulty and 97% of students who do not, with false positive and false negative rates of 6% and 46%, respectively.

When using the 25th percentile on the ORF-FP, 45% of students in Grade 6 with a reading difficulty and 90% of those who do not have a difficulty were identified, with a false positive rate of 13% and a false negative rate of 47%. In Grade 7, the sensitivity and specificity were 42% and 91%, respectively. In Grade 8, the sensitivity and specificity were 36% and 97%, respectively. As Table 3 shows, ORF-FP correctly identified a somewhat

greater proportion of struggling readers across Grades 6 to 8 than ORF-60 but slightly less adequate readers.

To further examine whether the classification accuracy of ORF-60 was more or less precise than ORF-FP, AUCs were also calculated for the ORF-60 and ORF-FP predicting group status on the TAKS Reading Assessment. In Grades 6 to 8, the AUC for ORF-60 predicting risk status on the TAKS was .78, .76, and .80, respectively. The AUC for ORF-FP predicting risk status on the TAKS was .78, .77, and .80 in Grades 6 to 8, respectively.

Discussion

Our first research question addressed whether duration of fluency assessment (first 60 s vs. full passage) substantively affected the reliability or validity of the assessment among middle-grade typically developing readers and middle-grade struggling readers.

Reliability

For reliability, results indicated that alternate form reliabilities of ORF-60 and ORF-FP were high for typically developing readers and struggling readers across Grades 6 to 8. In addition, the reliability of ORF calculated for the first 60 s was not substantively different than ORF calculated for the full passage among struggling readers and typically developing readers. As such, differences in reliability across grades would not explain differences in validity. Although the reliabilities reported in this study were lower than .90 reported by Espin et al. (2010) and Ticha et al. (2009), the general findings are similar. That is, regardless of duration, ORF measures were highly reliable for typically developing readers and struggling readers.

Validity

Examination of the concurrent validity coefficients suggested that although various trends in the data were observed, the validity of ORF-60 and ORF-FP was not substantively different. ORF-60 and ORF-FP correlated moderately well with external measures of reading fluency (i.e., AIMSweb Maze CBM Reading Comprehension, TOWRE, and TOSREC) across Grades 6 to 8 in typically developing and struggling readers. A trend observed across Grades 6 to 8 was that the magnitude of the relation between ORF-60 and ORF-FP and AIMSweb Maze CBM Reading Comprehension tended to be lower among struggling readers than typically developing readers, whereas the relation among ORF-60 and ORF-FP and TOWRE tended to be higher among struggling readers compared with typically developing readers.

Both methods of measurement were moderately correlated with GRADE Passage Comprehension and WJ-III Passage Comprehension across Grades 6 to 8 in both reader groups. Across Grades 6 to 8, the magnitude of the relation among ORF-60, ORF-FP, and GRADE Passage Comprehension tended to be lower among struggling readers compared with typically developing readers. This pattern of correlations may suggest that typically developing readers have attained a necessary threshold for word reading accuracy and fluency that allows them sufficient access to the meaning of print. In contrast, the magnitude of the relation among ORF-60, ORF-FP, and WJ-III Passage Comprehension tended to be

higher among struggling readers. The tendency for stronger relations with WJ-III Passage Comprehension among struggling readers is likely due to restriction of range.

Reliability and Validity

We hypothesized that among middle-grade readers, the influence of reading duration on ORF rates would vary by grade and reader skill. We hypothesized that among struggling, middle-grade readers, who may have not attained proficient levels of word reading accuracy, the most reliable and valid index of ORF would be measured within the first minute. In contrast, we hypothesized that among skilled readers who are actively engaging language comprehension processes to integrate information in text and information in text with background knowledge; ORF measured for the full passage would more accurately quantify their ability to fluently read for meaning and would be reflected in higher reliability and validity coefficients.

These hypotheses were not supported. Both struggling readers and typically developing readers possessed adequate speeded decoding skills for the two types of ORF tasks (reading for 60 s and the full passage) across Grades 6 to 8. Speeded decoding skills for struggling readers were on average one-half standard deviation below the mean; speeded decoding skills for typically developing readers were approximately one-third standard deviation above the mean. With both subgroups of readers possessing adequate speeded decoding skills, reading for longer duration provided both subgroups an opportunity to more fully engage language comprehension abilities (i.e., word and world knowledge) relative to reading for 60 s. However, reading for longer duration did not result in higher reliability or validity coefficients for either struggling readers or typically developing readers compared with reading for 60 s. For this reason, ORF for longer durations may be of limited practical value to classroom teachers given that it is less efficient to administer and provides potentially similar information relative to reading for 60 s.

Classification Accuracy

For the second research question, the classification results across Grades 6 to 8 suggested that when using the 25th percentile ORF-FP as the cut-off for prediction of reading risk on TAKS Reading Assessment, ORF-FP not only tended to correctly classify a greater proportion of struggling readers as at-risk but also tended to misclassify a slightly greater number of typically developing readers as at-risk. Across Grades 6 to 8, ORF-60 not only correctly classified a greater proportion of typically developing readers as not-at-risk but also misclassified a slightly greater proportion of at-risk readers as skilled. However, oversampling of struggling readers may have biased the positive and negative predictive values for both ORF-60 and ORF-FP. AUCs calculated to quantify the extent to which ORF-60 and ORF-FP correctly classified typically developing and struggling readers were not substantively different across grades and ranged from 0.76 to 0.80. Altogether, ORF calculated for the full passage was slightly more sensitive at identifying at-risk readers than ORF calculated for the first 60 s, but the sensitivity differences between ORF-FP and ORF-60 were small. Because the differences in sensitivity were small, one must weigh the benefits of identifying a higher percentage of at-risk students with reading difficulties against the added cost of administration time required for ORF-FP.

We hypothesized that if there were no major differences in reliability and validity, classification accuracy would be similar for both passage durations and both struggling and typically developing readers. This hypothesis was supported. No major differences in reliability and validity were observed and classification accuracy was similar for both durations of measuring ORF. We also hypothesized that the classification accuracy of ORF would be sufficiently accurate to provide middle schools with a method of differentiating at-risk and not-at-risk readers that parallels the proficient/not-proficient designations obtained from state accountability reading assessments. This hypothesis was moderately supported with AUCs for both durations (ORF-60 and ORF-FP) ranging from .76 to .80.

Regardless of duration, using a fluency score as the sole indicator of reading risk represents an imprecise method for identifying middle-grade at-risk students because of the measurement error associated with the assessment (Shepard, 1980). Because of measurement error, any attempt to set a cut-point for the purpose of distinguishing risk status, tends to lead to classification errors because scores fluctuate around the cut-point (Francis et al., 2005). Score fluctuation results because no single score can perfectly capture a student's true reading ability and increases as the cut-point moves away from the center of the ability distribution (because most reading tests are designed to yield maximal precision in the center of the score distribution; Francis et al., 2005). Thus, single assessments that also use a cut-point at the lower end of the score distribution are highly likely to miss a number of students who would benefit from supplemental reading intervention and misclassify adequate readers as at-risk.

Ways of Improving Classification Accuracy

One method of addressing measurement error associated with using a single score from a single assessment to index reading ability is to use a method of identification that uses multiple measures, which may result in a higher percentage of at-risk students receiving preventive or remedial intervention services as needed (Barth et al., 2008; Fletcher et al., 2011; Johnson, Jenkins, Petscher, & Catts, 2009). For example, Compton, Fuchs, Fuchs, and Bryant (2006); Davis, Lindo, and Compton (2007), Jenkins, Hudson, and Johnson (2007); and Riedel (2007) suggested that a screening battery that included a measure of ORF and a measure of vocabulary or comprehension would increase correct identification of students at-risk for reading failure. Johnson et al. (2009) reported that combining ORF with a measure of vocabulary resulted in a 2% increase in specificity among first-grade readers. Fletcher et al. (2011) found better coverage of inadequate responders to a Tier 2 reading intervention when multiple measures of fluency were used. It is also possible that classification accuracy can be further increased by incorporating professional judgment or using multiple years of data (Gilbert, Compton, Fuchs, & Fuchs, 2012).

A school-based approach that takes advantage of multiple sources of information would likely lead to improved identification and remediation of middle school students with reading disabilities and reading difficulties. Such an approach might be a multiple-gating method where data from state accountability tests of reading comprehension are used to identify a reading-impaired pool of students. Students in this pool are then tested using an ORF assessment, which only takes 1 min to administer. Data from the ORF assessment

could be used to triage students into two different types of reading classes. Students with low ORF rates would participate in a reading class that places heavy emphasis on word study and reading fluency and their connection to reading comprehension. Students with adequate ORF rates would participate in a class that places heavy emphasis on reading comprehension and the acquisition of content area knowledge. In addition, other data that might explain inconsistent performance on high-stakes reading tests could be examined, such as behavior and attention and other academic areas such as spelling and writing. This type of multiple-gating approach would help to explain why students fail high-stakes reading tests and how instruction could be designed to address those deficits at the middle school level. Finally, the use of a multiple-gating approach would reduce false-identification rates (false negatives and false positives) thereby resulting in more accurate identification of students' reading abilities and more efficient allocation of educational resources.

Limitations

The sensitivity and specificity rates used in this study are specific to the sample, measures, and cut-points we adopted. In particular, the cut-point of the 25th percentile used in this study, while commonly used for identification, is arbitrary. Moving the cut-point lower would likely decrease sensitivity and increase specificity, while moving it higher would have the opposite effect. False positives and false negatives are always a trade-off related in part to the cut-point (Glover & Albers, 2007).

Another potential limitation of this study is that it did not formally examine where on the continuous distribution of ORF scores duration has the greatest impact. By looking at the interaction of duration and rate, the area of the distribution most impacted by duration might be better elucidated. By understanding the area of the distribution more likely to be influenced by duration, one can then select the best score to index ORF ability.

This study also did not measure growth over time and examine whether duration differentially impacts the reliability of the slope estimate for different subgroups of readers. Because progress monitoring takes time away from classroom instruction, the information obtained from the assessment should be useful to classroom teachers (Glover & Albers, 2007). Future research should more formally examine whether reading duration impacts the reliability of the slope estimate and whether it varies by reader subgroup, so that classroom teachers can determine whether the information obtained is more useful, practical, and relevant than what is obtained in the first minute of reading (Glover & Albers, 2007).

The last limitation of this study was that the GRADE Passage Comprehension standard score was prorated because testing constraints imposed by the schools did not permit administration of both the Passage Comprehension subtest and Sentence Comprehension subtest to the full sample of middle-grade readers. The prorated standard score represents the proportion of items correct divided by the items administered. As a consequence, correlations among measures of ORF with the GRADE Passage Comprehension prorated score should be replicated in other samples of middle-grade readers to provide additional support for the finding that magnitude of these correlations is moderate among struggling and typically developing readers.

Implications for Practice

A substantial body of literature reports high reliability and validity coefficients for ORF probes among elementary-grade readers and middle-grade readers (Reschly, Busch, Betts, Deno, & Long, 2009; Ticha et al., 2009; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). Despite this research, gaps are present in the literature base. There is growing interest in the use of ORF probes among middle-grade struggling readers and students in special education for the purpose of monitoring reading progress and instructional decision making (Ticha et al., 2009; Vaughn, Cirino, et al., 2010). For example, in the state of Texas, middle school teachers are advised to administer this particular measure 3 times a year (i.e., beginning of year, middle of year, and end of year) to more accurately place students in Tier 2 interventions that focus on word reading, fluency, and comprehension, fluency and comprehension, or comprehension only. The need for progress monitoring tools is also essential for students with learning disabilities, whose reading progress can be slow and incremental without implementation of highly effective interventions and whose reading progress may not be captured by end of year final status assessments (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Ticha et al., 2009). This study attempted to fill this gap in the literature by demonstrating that ORF probes for the first minute of reading or for the full passage are reliable and valid measures of indexing reading ability among struggling middle-grade readers and may be used as part of a secondary-school teachers' intervention design to measure reading abilities and inform instructional decisions.

Acknowledgments

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported, in part, by Grants P50 HD052117 and K08 HD068545-01A1 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) that were awarded to the University of Houston as well as Grant R305F100013—Reading for Understanding from the Institute of Education Sciences, U.S. Department of Education that was awarded to the University of Texas at Austin.

References

- Ardoin SP, Christ TJ. Curriculum based measurement of oral reading: Estimates of standard error when monitoring progress using alternate passage sets. *School Psychology Review*. 2009; 38:266–283.
- Barth AE, Stuebing K, Anthony J, Denton C, Mathes P, Fletcher JM, Francis D. Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences*. 2008; 18:296–307. [PubMed: 19081758]
- Brasseur-Hock I, Hock M, Kieffer M, Biancarosa G, Deshler D. Adolescent struggling readers in urban schools: Results of a latent class analysis. *Learning and Individual Differences*. 2011; 21:438–452.
- Cirino P, Romain M, Barth AE, Tolar T, Fletcher JM, Vaughn S. Reading skill components and impairments in middle school struggling readers. *Reading and Writing*. 2013; 26:1059–1086. [PubMed: 24000271]
- Compton DL, Fuchs D, Fuchs LS, Bryant JD. Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*. 2006; 98:394–409.

- Crawford L, Tindal G, Stieber S. Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment*. 2001; 7:303–323.
- Daane, MC.; Campbell, JR.; Grigg, WS.; Goodman, MJ.; Oranje, A. Fourth-grade students reading aloud: NAEP 2002 special study of oral reading (NCES 2006-469). Washington, DC: Institute of Education Sciences, National Center for Education Statistics, U.S. Department of Education; 2005.
- Davis GN, Lindo EJ, Compton D. Children at-risk for reading failure: Constructing an early screening measure. *TEACHING Exceptional Children*. 2007; 39:32–39.
- Deno SL, Fuchs LS, Marston D, Shin J. Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*. 2001; 30:507–526.
- Espin CA, Deno SL. Content-specific and general reading disabilities of secondary-level students: Identification and educational relevance. *The Journal of Special Education*. 1993a; 27:321–337.
- Espin CA, Deno SL. Performance in reading from content area text as an indicator of achievement. *Remedial and Special Education*. 1993b; 14:47–59.
- Espin CA, Foegen A. Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children*. 1996; 62:497–514.
- Espin CA, Wallace T, Lembke E, Campbell H, Long JD. Creating a progress-monitoring system in reading for middle-school students: Tracking progress toward meeting high-stakes standards. *Learning Disabilities Research & Practice*. 2010; 25:60–75.
- Fletcher JM, Stuebing KK, Barth AE, Denton CA, Cirino PT, Francis DJ, Vaughn S. Cognitive correlates of inadequate response to intervention. *School Psychology Review*. 2011; 40:2–22.
- Francis, DJ.; Barth, A.; Cirino, P.; Reed, D.; Fletcher, J. The Texas Middle School Fluency Assessment. Austin: Texas Educational Agency; 2008.
- Francis DJ, Fletcher JM, Stuebing K, Lyon R, Shaywitz B, Shaywitz S. Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities*. 2005; 38:98–108. [PubMed: 15813593]
- Francis DJ, Santi KL, Barr C, Fletcher JM, Varisco A, Foorman BR. Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology*. 2008; 46:315–342. [PubMed: 19083362]
- Fuchs LS, Fuchs D, Hamlett CL, Walz L, Germann G. Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review*. 1993; 22:27–48.
- Gilbert JK, Compton DL, Fuchs D, Fuchs LS. Early screening for risk of reading disabilities: Recommendations for a four-step screening system. *Assessment for Effective Intervention*. 2012; 38:6–14. [PubMed: 24478613]
- Glover T, Albers C. Considerations for evaluation universal screening assessments. *Journal of School Psychology*. 2007; 45:117–135.
- Hasbrouck J, Tindal GA. Oral reading fluency norms: A valuable assessment tool for reading teachers. *Reading Teacher*. 2006; 59:636–644.
- Hintze JM, Silbergliitt B. A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*. 2005; 34:372–386.
- Hock MF, Brasseur IF, Deshler DD, Catts HW, Marquis JG, Mark CA, Wu Stribling J. What is the reading component skill profile of adolescent struggling readers in urban schools? *Learning Disability Quarterly*. 2009; 32:21–38.
- Jenkins JR, Hudson RF, Johnson ES. Screening for service delivery in an RTI framework: Candidate measures. *School Psychology Review*. 2007; 36:582–599.
- Johnson ES, Jenkins JR, Petscher Y, Catts HW. How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice*. 2009; 24:174–194.
- Lexile Framework. Lexile Framework for Reading [Computer Software]. Durham, NC: MetaMetrics; 2007.
- McGlinchey MT, Hixson MD. Using curriculum based measurement to predict performance on state assessments in reading. *School Psychology Review*. 2004; 33:193–203.
- McGrew, KS.; Woodcock, RW. Woodcock-Johnson III technical manual. Itasca, IL: Riverside; 2001.
- Nolet, V.; McLaughlin, MJ. Accessing the general curriculum: Including students with disabilities in standards-based reform. Thousand Oaks, CA: Corwin Press; 2000.

- Reschly AL, Busch TW, Betts J, Deno ST, Long JD. Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*. 2009; 47:427–469. [PubMed: 19808123]
- Riedel BW. The relation between DIBELS, reading comprehension, and vocabulary among urban first-grade students. *Reading Research Quarterly*. 2007; 42:546–567.
- Shepard L. An evaluation of the regression discrepancy method for identifying children with learning disabilities. *The Journal of Special Education*. 1980; 14:79–91.
- Shinn, MR.; Shinn, MM. AIMSweb training workbook: Administration and scoring of reading Maze for use in general outcome measurement. Eden Prairie, MN: Edformation; 2002.
- Silbergliitt B, Burns MK, Madyun NH, Lail KE. Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools*. 2006; 43:527–535.
- Silbergliitt B, Hintze J. Formative assessment using CBM-R cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment*. 2005; 23:304–325.
- Texas Educational Agency. Information booklet: Reading, Grade 5–Revised. Author; 2004. TAKS: Texas Assessment of Knowledge and Skills. Retrieved from www.tea.state.tx.us/index3.aspx?id=3693&menu_id=793
- Ticha R, Espin C, Wayman MW. Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading aloud and maze selection measures. *Learning Disabilities Research & Practice*. 2009; 24:132–142.
- Torgesen, J.; Nettles, S.; Howard, P.; Winterbottom, R. Technical Report No 6. Tallahassee: Florida Center for Reading Research; 2005. Brief report of a study to investigate the relationship between several brief measures of reading fluency and performance on the Florida Comprehensive Assessment Test-Reading in 4th, 6th, 8th, and 10th grades.
- Torgesen, J.; Wagner, R.; Rashotte, C. Test of Word Reading Efficiency. Austin, TX: Pro-Ed; 1998.
- Vaughn S, Cirino PT, Wanzek J, Wexler J, Fletcher JM, Denton CD, Francis DJ. Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review*. 2010; 39:3–21. [PubMed: 21479079]
- Vaughn S, Wanzek J, Wexler J, Barth A, Cirino PT, Fletcher JM, Francis DJ. The relative effects of group size on reading progress of older students with reading difficulties. *Reading and Writing*. 2010; 23:931–956. [PubMed: 21072131]
- Wagner, R.; Torgesen, J.; Rashotte, C.; Pearson, N. Test of Silent Reading Efficiency and Comprehension. Austin, TX: Pro-Ed; 2010.
- Wallace T, Espin C, McMaster K, Deno S, Foegen A. CBM progress monitoring within a standards-based system: Introduction to the special series. *The Journal of Special Education*. 2007; 41:66–67.
- Wayman MM, Wallace T, Wiley HI, Ticha R, Espin CA. Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*. 2007; 41:85–120.
- Williams, KT. Group Reading Assessment Diagnostic Evaluation. Shoreview, MN: Pearson AGS Globe; 2001.
- Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III Test of Cognitive Abilities. Itasca, IL: Riverside; 2001.
- Yovanoff P, Duesbery L, Alonzo J, Tindal G. Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice*. 2005; 24:4–12.

Table 1

Demographics by Grade and Group ($N = 1,474$ Students).

Variable	Grade 6			Grade 7			Grade 8		
	Typically developing readers	Struggling readers		Typically developing readers	Struggling readers		Typically developing readers	Struggling readers	
<i>n</i>	228	337		160	190		245	312	
Age									
<i>M</i>	11.23	11.56		12.19	12.46		13.13	13.35	
<i>SD</i>	0.48	0.68		0.46	0.68		0.35	0.58	
% Male	40	48		43	54		42	57	
% Free or reduced lunch	46	78		48	72		49	72	
% English as a second language	1	7		3	20		0	21	
% African American	40	47		43	33		36	39	
% White	29	12		24	14		29	16	
% Hispanic	27	40		26	50		30	42	
% Other	4	1		7	3		5	3	

Table 2

Means and Standard Deviations for ORF-60, ORF-FP, and External Measures of ORF and Reading Comprehension by Reader Subgroup Within Grade ($N = 1,474$ Students).

Variable	Grade 6		Grade 7		Grade 8	
	Typical readers ($n = 228$)	Struggling readers ($n = 337$)	Typical readers ($n = 160$)	Struggling readers ($n = 190$)	Typical readers ($n = 245$)	Struggling readers ($n = 312$)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
ORF-60	138.7 (30.4)	105.7 (30.7)	143.8 (30.6)	111.1 (33.6)	154.4 (26.8)	119.3 (32.8)
ORF-FP	135.2 (30.9)	102.2 (29.8)	141.5 (31.1)	107.5 (32.8)	151.9 (27.5)	116.4 (32.9)
AIMSweb Maze	188.8 (61.9)	143.6 (52.8)	196.6 (63.0)	159.5 (64.9)	210.0 (59.7)	163.5 (61.2)
TOSREC	99.9 (12.4)	86.1 (11.1)	100.3 (12.1)	83.6 (12.9)	97.9 (14.7)	79.4 (13.4)
TOWRE	105.1 (14.4)	92.5 (14.3)	104.4 (13.1)	92.5 (15.9)	104.8 (12.8)	91.3 (14.6)
GRADE Passage Comprehension	101.6 (11.4)	88.9 (9.1)	100.8 (11.9)	86.2 (10.2)	104.0 (12.5)	87.8 (10.8)
WJ-III Passage Comprehension	99.4 (10.2)	87.0 (11.0)	98.2 (8.9)	85.4 (11.5)	100.1 (9.4)	85.6 (10.9)

Note. ORF-60 scores represent the average words read correctly for the first minute of reading for five passages; ORF-FP scores represent the average words read correctly per minute for the full passage for five passages; AIMSweb Maze scores represent the total number of words plus targets read in 3 min; TOSREC, TOWRE, GRADE Passage Comprehension, and WJ-III Passage Comprehension scores represent standard scores. ORF = oral reading fluency; ORF-FP = oral reading fluency calculated for the full passage; AIMSweb Maze = AIMSweb Maze Reading Comprehension assessment; TOSREC = *Test of Silent Reading Efficiency and Comprehension*; TOWRE = *Test of Word Reading Efficiency*, composite score; GRADE Passage Comprehension = *Group Reading Assessment and Diagnostic Evaluation*, Passage Comprehension subtest; WJ-III Passage Comprehension = *Woodcock Johnson-III* Passage Comprehension subtest.

Table 3

Classification Accuracy Statistics for ORF-60 and ORF-FP Predicting Reader Classification on the TAKS Reading Assessment in Grades 6 to 8 ($N = 1,474$ Students).

Measure and benchmark	Sensitivity	Specificity	False positive	False negative	Positive predictive rate	Negative predictive rate	Overall classification accuracy
Grade 6							
25th% ORF-60	.40	.92	.12	.49	.88	.51	.61
25th% ORF-FP	.45	.90	.13	.47	.87	.53	.63
Grade 7							
25th% ORF-60	.38	.94	.12	.44	.88	.56	.64
25th% ORF-FP	.42	.91	.15	.43	.85	.57	.65
Grade 8							
25th% ORF-60	.36	.97	.06	.46	.94	.54	.63
25th% ORF-FP	.40	.94	.10	.45	.90	.55	.64

Note. ORF-60 scores represent the average words read correctly for the first minute of reading for five passages. ORF-FP scores represent the average words read correctly per minute for the full passage for five passages. ORF = oral reading fluency; ORF-FP = ORF calculated for the full passage; TAKS = *Texas Assessment of Knowledge and Skills*.