

The Relations Among Oral and Silent Reading Fluency and Comprehension in Middle School: Implications for Identification and Instruction of Students With Reading Difficulties

Carolyn A. Denton

*Children's Learning Institute, University of Texas Health Science Center
Houston*

Amy E. Barth and Jack M. Fletcher

University of Houston

Jade Wexler and Sharon Vaughn

University of Texas at Austin

Paul T. Cirino, Melissa Romain, and David J. Francis

University of Houston

The purpose of this study was to investigate the relations among oral and silent reading fluency and reading comprehension for students in Grades 6 to 8 ($n = 1,421$) and the use of fluency scores to identify middle school students who are at risk for failure on a high-stakes reading test. Results indicated moderate positive relations between measures of fluency and comprehension. Oral reading fluency (ORF) on passages was more strongly related to reading comprehension than ORF on word lists. A group-administered silent reading sentence verification test approximated the classification accuracy of individually administered ORF passages. The correlation between a maze task and comprehension was weaker than has been reported for

elementary students. The best predictor of a high-stakes reading comprehension test was the previous year's administration of the grade-appropriate test; fluency and verbal knowledge measures accounted for only small amounts of unique variance beyond that accounted for by the previous year's administration.

In recent years, increasing emphasis has been placed on identifying and intervening with older students with reading difficulties (e.g., Biancarosa & Snow, 2004), but questions remain regarding the valid identification of these students. Based on its demonstrably strong relations with both decoding and reading comprehension in the elementary grades, the measurement of oral reading fluency (ORF) has been widely accepted as an efficient and valid way to assess elementary school students' overall reading competence (e.g., Fuchs, Fuchs, Hosp, & Jenkins, 2001). These measures are often administered to elementary school students to identify those who need reading intervention, to monitor their progress, and to determine whether they have responded adequately to instruction.

The empirical support for the use of such measures with elementary grade students is strong, but there is appreciably less known about valid applications of fluency measures with older students. Assumptions have been made that the measures will operate with secondary school students as they do with younger children, but these assumptions have not been substantiated. The purpose of this study was to examine (a) the relations among multiple measures of oral and silent reading fluency and reading comprehension for students in Grades 6, 7, and 8 and (b) the use of fluency measures to identify students at risk for failure on a high stakes reading comprehension test who may need supplemental reading intervention.

MEASURING ORF

ORF is most often measured by timing a student reading one or more passages orally for 1 to 2 min, counting the total number of words read during that time period, subtracting error words, and calculating the number of words read correctly per minute (e.g., Hasbrouck & Tindall, 2006). In some applications, oral passage reading fluency is measured using reading passages actually taken from the reading curriculum used in the students' classrooms (Fuchs & Deno, 1994). In others, standardized passages resembling those in grade-level reading programs are used (e.g., Good & Kaminski, 2002). ORF can also be measured by having students read lists of words in isolation rather than connected text. Fuchs et al. (2001) summarized research indicating that word list reading fluency and text reading fluency may be measuring somewhat different constructs. Based on the observation that text fluency may be more strongly related to reading comprehension than

word list fluency, Fuchs et al. suggested that word list fluency represents students' word recognition proficiency, whereas passage reading fluency represents how efficiently they process information beyond the word level.

MEASURING SILENT READING FLUENCY

Because of the time required to individually administer tests of ORF, and because older students primarily read silently rather than orally, assessments of silent reading fluency may be useful for identifying older students with reading difficulties and monitoring their progress. The measurement of silent reading fluency presents obvious challenges because silent reading is not an observable behavior. Some researchers have used student self-report to determine how many words students can read silently in a specific period. For example, Fuchs et al. (2001) described an unpublished analysis of data (Fuchs, Fuchs, Eaton, & Hamlett, 2000, as cited by Fuchs et al., 2001) obtained when fourth-grade students read a passage silently and circled the last word read when told to stop after 2 min. Students answered comprehension questions following silently and orally read passages and took a standardized comprehension test. The researchers report significantly lower correlations for silent reading than for oral reading with both comprehension measures and speculated that students may have inaccurately reported the last word read during silent reading.

Two standardized measures of silent reading fluency take a different approach, presenting students with strings of words with no spaces separating them; the task is to draw lines separating the words, and the score is the number of words correctly identified in three minutes. In the Test of Silent Word Reading Fluency (TOSWRF; Mather, Hammill, Allen, & Roberts, 2004), students are presented with a string of unrelated words without spaces. For example, when presented with *istworunsawuse*, the student would separate it into words as follows: */is/two/run/saw/use/*. The format of the Test of Silent Contextual Reading Fluency (TOSCRF; Hammill, Wiederholt, & Allen, 2006) is similar, except that students are presented with text passages with all words printed in uppercase letters with no separations between words and no punctuation or spaces between sentences; again, the task is to draw lines to separate words.

Other approaches used to measure silent reading fluency are sentence verification and maze tasks. These tests may directly measure both fluency and comprehension, but for purposes of this study we consider them measures of silent reading fluency. In a sentence verification test, students read sentences and must indicate whether they are true or false. The items are designed so that, if the sentences are read correctly, they can be identified as true or false using general world knowledge (e.g., All birds are blue.). The score is the number of correct answers obtained within a given period. A widely used example of a sentence verification

measure is the Woodcock–Johnson III (WJ III; Woodcock, McGrew, & Mather, 2001) Reading Fluency subtest.

A maze is a multiple-choice close test. In a maze, words are systematically omitted from a passage (usually every seventh word), and each is replaced by three word choices (the original word, a near distracter, and a far distracter). The task is to select from the three choices the one word that appeared in the original passage. Maze tests are often timed for 3 min.

The relation between ORF and silent reading fluency is complicated by the fact that students may exhibit different levels of comprehension when they read orally and silently (Miller & Smith, 1990). Although the research in this area is somewhat ambiguous, there is evidence that this effect on comprehension may vary according to reading level (e.g., Miller & Smith, 1990). On the other hand, Hale et al. (2007) reported that students in Grades 4, 5, 10, 11, and 12 were better able to answer comprehension questions following oral reading than after reading silently and that this was equally true across grade levels. It is difficult to know whether differences detected in students' oral and silent reading comprehension are due to the way silent reading fluency is measured or to real differences in understanding and remembering what is read.

READING FLUENCY IN THE ELEMENTARY GRADES

Researchers consistently document a moderate to strong positive relationship between ORF and reading comprehension for students in the elementary grades (Fuchs et al., 2001; Marston, 1989; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). For example, Hosp and Fuchs (2005) reported correlations of .79 to .84 between ORF and the Passage Comprehension subtest of the Woodcock Reading Mastery Test–Revised in Grades 1 to 4. The maze silent reading task has also been shown to be positively related to reading comprehension for students in the elementary grades; however, ORF appears more strongly related to comprehension than maze tasks for elementary grade students (Ardoin, Witt, & Suldo, 2004; Jenkins & Jewell, 1993).

Identification of Reading Risk

There is evidence that measures of ORF are useful for screening elementary-grade students to identify those at risk for failing high stakes reading comprehension tests (Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Schilling, Carlisle, Scott, & Zheng, 2007; Vander Meer, Lentz, & Stollar, 2005). The use of ORF measures alone for this purpose appears to result in moderate rates of inaccurate prediction. Both McGlinchey and Hixson (2004) and Stage and Jacobsen (2001) reported that ORF correctly classified about 75% of participating

fourth-grade students as passing high-stakes tests, meaning that about one fourth of the students in these studies were incorrectly classified. The consequences of these types of errors are that students who are in need of intervention may be deprived of that assistance and/or school resources may be used to provide intervention to students who do not need it.

When classification accuracy is evaluated, the percentages of students falsely identified as at risk or not at risk are determined by the cut score or benchmark selected for the prediction assessments. For example, if the ORF benchmark is liberal, many students will be classified as at risk; some of these will be truly at risk (true positives), and some will be identified in error (false positives). On the other hand, if the ORF benchmark is set at a fairly low level, many students will meet that benchmark and thus be classified as not at risk; some will be truly not at risk (true negatives) when some are actually in need of intervention (false negatives). As the consequence of false negative errors is that students in need of intervention do not receive it, these are usually considered more serious. Some propose that false negative errors be kept below 10% so that students in need of intervention are not deprived of it (e.g., Jenkins, 2003).

READING FLUENCY IN GRADES 6 TO 8

Few studies have investigated the relation of fluency and comprehension for students at the secondary level, with most researchers reporting lower correlations than commonly reported for elementary school students. An exception is Fuchs, Fuchs, and Maxwell (1988), who reported a correlation of .91 between ORF and standardized test reading comprehension scores for middle school students with reading disabilities.

The relation between ORF and comprehension may weaken as students progress to higher grade levels. Jenkins and Jewell (1993) documented this pattern for students in Grades 2 through 6, as did Shinn, Knutson, Collins, Good, and Tilly (1992) for Grades 3 and 5. Similarly, Silbergitt, Burns, Madyun, and Lail (2006) found that correlations between ORF and a state accountability test were significantly lower at Grades 5, 7, and 8 than at Grade 3 and that the relation was significantly weaker at Grade 8 than at Grade 5. Moreover, in this study ORF accounted for 50.4% of the variance in comprehension test scores at Grade 3 but only 26.0% at Grade 8. Schatschneider et al. (2004) similarly found that not only text fluency but also reasoning and verbal knowledge accounted for comparable variance on the Florida Comprehensive Assessment Test (FCAT) in Grades 7 and 10, whereas text fluency accounted for substantially more variance in third grade.

Conversely, Torgesen, Nettles, Howard, and Winterbottom (2003) reported that correlations between ORF and comprehension, as measured by the FCAT, were fairly stable across Grades 4 through 10 but were lower than those often reported

in the primary grades. Torgesen et al. also investigated the relation of silent reading fluency measures and the FCAT, with mixed results. Aligned with their findings for ORF, correlations between the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010), a sentence verification task, and the FCAT were similar across grade levels. Relations between the TOSCRF and FCAT declined from Grade 4 to Grades 8 and 10 and were significantly lower than those for any other measures. Scores on a maze task were more highly correlated with FCAT scores in Grade 8 than in Grade 4, although the correlation declined sharply at Grade 10.

Identification of Reading Risk

Little research has specifically evaluated the use of reading fluency measures to identify secondary school students in need of reading intervention. Fewster and Macmillan (2002) reported that ORF data collected in Grades 6 and 7 were significant predictors of the grades students received in English and social studies classes in Grades 8, 9, and 10 and accurately differentiated between students who attended special education, remedial, general education, and honors classes.

In general, for middle school students, vocabulary tasks may be better predictors of comprehension than timed oral reading. For example, Yovanoff, Duesbery, Alonzo, and Tindal (2005) examined the relative contributions of vocabulary and fluency to comprehension (measured by answering questions following passage reading) across Grades 4 to 8. They found that both vocabulary and fluency were significant predictors of comprehension at all grade levels but that the relative importance of ORF decreased in the higher grades. Similarly, Espin and Foegen (1996) reported that brief timed measures requiring middle school students to match vocabulary words with their definitions were better predictors of student performance on content-area reading tasks (e.g., answering questions after reading expository text) than maze or ORF tasks. In fact, when entered into a multiple regression, ORF and maze accounted for little or no additional variance beyond the vocabulary task.

PURPOSE AND RESEARCH QUESTIONS

Existing research on the relation of reading fluency and comprehension at the middle school level is limited and has yielded mixed results. Much of this research indicates that the relation of ORF and comprehension declines as students progress from the primary to the intermediate and secondary grades, when they are faced with more complex text and greater demands for high-level reasoning and inferencing (Paris, Carpenter, Paris, & Hamilton, 2005). Even less is known about the relation between silent reading fluency and comprehension

at the secondary level. Moreover, research on the relation between fluency and comprehension has been influenced by the variety of tests that have been used to measure these domains. Our review indicated a need for research validating the use of fluency measures to identify middle school students in need of reading intervention. Using multiple measures of oral and silent reading fluency and reading comprehension, we addressed the following research questions:

1. What are the relations among ORF and silent reading fluency, verbal knowledge, and reading comprehension in Grades 6, 7, and 8? Do these relations differ when reading fluency and reading comprehension are measured in different ways? Specifically, what are the relations with reading comprehension when oral fluency is measured using word lists or passages and when silent fluency is measured using maze, word identification, or sentence verification tasks?
2. How much unique variance in state reading test scores is explained by measures of reading fluency and verbal knowledge after accounting for the previous year's performance on the state reading test?
3. How accurate are ORF passage fluency tests, silent reading fluency test, and the previous year's high-stakes reading comprehension test for the identification of middle school readers who need intervention because they are at risk for failure on a subsequent high-stakes test?

METHOD

Participants

School sites. Participants were selected from seven middle schools (serving Grades 6–8) located in the southwestern United States. Three of the study schools were from a large urban district in one city, and four schools were from two school districts located near a smaller city. School populations ranged from 633 to 1,300 students. Students qualifying for reduced or free lunch ranged from 56% to 86% across the schools in the urban site and from 40% to 85% in the small city site. At the inception of the study, all schools were rated “acceptable” or higher according to state standards.

Students. Participants in this study were 1,421 sixth-, seventh-, and eighth-grade students selected during the 2006–2007 academic year. The sample included 564 students in Grade 6, 312 students in Grade 7, and 545 students in Grade 8. The sample was ethnically diverse, with 39% of participants African American, 38% Hispanic, 19% Caucasian, and 4% Asian and Other; 48% were male. Of the 1,421 students, 764 (54%) were classified as struggling readers and

655 (46%) as typical readers. Struggling readers were defined as students who (a) failed the state reading achievement test, the Texas Assessment of Knowledge and Skills (TAKS; $n = 528$) or (b) performed within one half of one standard error of measurement above the pass–fail cut point on TAKS–Reading on their first attempt in the spring of 2006 (i.e., scale scores ranging 2,100–2,150 points; $n = 236$). We include students who scored within the test’s standard error to account for the inherent imprecision of using single test scores to classify students. Typical readers were defined as students who scored above a scale score of 2,150 on the TAKS–Reading (i.e., greater than one half of one standard error of measurement above the pass–fail cut point on the TAKS). Because a large proportion of students passed the TAKS, we randomly selected typical readers within school (and grade) in proportion to the number of struggling readers. Struggling readers were oversampled so that the results of this study could be more readily generalized to the population of middle school students most likely to be administered measures of reading fluency. Even though struggling readers were oversampled, all variables used in the analyses were normally distributed. This reflects the heterogeneity of the reading deficits present among struggling readers. That is, struggling readers did not fail TAKS–Reading for one specific reason; rather struggling readers failed because of deficits in the areas of decoding, reading fluency, and/or reading comprehension. Students were excluded from the study if (a) they were enrolled in a special education life skills class, had a sensory disorder, or severe disability; (b) their Modified TAKS–Reading performance levels were below a third grade level; or (c) they were enrolled in English as a Second Language classes.

Procedures

Participants were administered a comprehensive battery of reading assessments in the fall of 2006, including multiple measures of reading comprehension, ORF, silent reading fluency, and vocabulary. In the spring of 2007 they were administered the reading subtest of the TAKS. A detailed description of the measures can be found at <http://www.texasldcenter.org/outcomes>. For all measures except the TAKS, participants were assessed by trained examiners, each of whom had demonstrated at least 95% accuracy during practice assessments. The TAKS was administered by students’ regular teachers, using standardized procedures.

Reading Comprehension Measures

TAKS (Texas Education Agency, 2004a, 2004b). The TAKS is the Texas academic accountability test. It is untimed, criterion referenced, and aligned with grade-level state curriculum standards. The TAKS reading subtest requires that students read expository and narrative texts and answer multiple-choice questions

that assess understanding of the literal meaning of the passages, vocabulary, and various aspects of critical reasoning. Standard scores are the dependent measure used in this study. For the 2006 and 2007 administrations, internal consistencies of the TAKS reading test across grades ranged from .87 to .89.

Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001). The Passage Comprehension subtest of the GRADE requires reading of passages with multiple-choice responses. The GRADE is a group-based, norm-referenced, untimed test of reading comprehension. Coefficient alpha for students in Grades 6 to 8 is .82 to .88. The dependent measure analyzed was a prorated standard score computed for Passage Comprehension.

WJ III Tests of Achievement, Passage Comprehension (Woodcock et al., 2001). The Passage Comprehension subtest is an individually administered, norm-referenced, untimed test of reading comprehension that utilizes a cloze procedure. The coefficient alpha for students in Grades 6 to 8 is .94 to .96. The dependent measure analyzed was the norm-referenced standard score.

Oral Reading Fluency Measures

ORF Curriculum-Based Measurement (CBM) Passage Fluency (University of Houston, 2008). The passage reading fluency task is an individually administered, timed assessment of oral fluency in connected text. The measure consists of graded expository or narrative passages administered as 1-min probes. All passages are approximately 500 words in length and range in difficulty from 350 to 1,400 Lexiles (Lennon & Burdick, 2004). The raw score is the linearly equated number of words read correctly within 1 min. Linearly equated scores were utilized to minimize the effects of text type, order of administration, and text difficulty in the estimation of students' oral reading fluency abilities (Francis et al., 2008). Coefficient alpha for students is .87 in Grade 6 and .96 for students in Grade 7–8 (Vaughn et al., 2010). Each student was administered five reading fluency passages. The dependent measure analyzed was an average linearly equated score for the five passage read.

Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999). On the Sight Word Efficiency subtest of the TOWRE the student is given a list of 104 increasingly challenging 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 sec. The test-retest reliability coefficients are at or above .90 for students in middle school. The standard score was the dependent measure utilized.

ORF CBM Word Fluency (University of Houston, 2008). Each student was administered three word lists from the Oral Reading Fluency CBM Word Fluency measure, an individually administered, timed assessment of word reading accuracy and fluency. For each of three word lists, students are required to read as many words as possible within 1 min. The raw score is the linearly equated number of words read correctly within 1 min for each word list. Linearly equated scores were utilized to control for form effects on the estimation of reading fluency abilities (Francis et al., 2008). The dependent measure utilized is the average linearly equated score for the three lists read. The mean intercorrelation of the three word lists read in the entire sample was 0.92 for students in Grade 6 and 0.97 for students in Grades 7 and 8.

Silent Reading Fluency Measures

AIMSweb Maze CBM (Shinn & Shinn, 2002). Each student was administered one AIMSweb Maze CBM passage. The Maze CBM is a 3-min, group-based assessment of silent reading fluency and comprehension. Howe and Shinn (2002) reported a median test-retest reliability of 0.85 for students in Grade 6, 0.79 for students in Grade 7, and 0.92 for students in Grade 8. The raw score is the number of targets correctly identified in 3 min and was the dependent measure utilized.

TOSREC (Wagner et al., 2010). The TOSREC is a 3-min, group-based sentence verification task designed to assess silent reading fluency and comprehension. The raw score is the number of correct minus the number of incorrect responses within the 3-min time limit. The mean intercorrelations across the five performances in the 1st year of the study was 0.79 for standard scores for students in Grade 6 and 0.92 for students in Grades 7 and 8. The standard score was the dependent measure utilized.

TOSCRF (Hammill et al., 2006). The TOSCRF, an assessment of silent reading fluency, measures the speed with which students can recognize the individual words in a series of passages, presented with no spaces between words and no punctuation, within a 3-min time limit. The average test-retest correlation for students in middle school is .84. The standard score was the dependent measure utilized.

Measure of Verbal Knowledge (Vocabulary)

Kaufman Brief Intelligence Test-2, Verbal Knowledge (KBIT-2; Kaufman & Kaufman, 2004). The Verbal Knowledge subtest is an individually administered, norm-referenced, untimed test of receptive vocabulary and

general information. Internal consistency values for the subtests and composite range from .87 to .95, and test–retest reliabilities range from .80 to .95, in the age range of the students in this study (Kaufman & Kaufman, 2004). The KBIT–2 also has a Riddles subtest that was not utilized; instead, the Verbal Knowledge score was prorated for the verbal domain, and therefore verbal standard scores were utilized.

RESULTS

Table 1 presents Pearson correlation coefficients illustrating the relations among fluency, verbal knowledge, and comprehension for students in Grades 6 to 8 (Question 1). It indicates that some fluency measures appear to correlate more strongly with comprehension than others and that these relations are fairly

TABLE 1
Correlation Matrix of Pretest Variables

Variable	1	2	3	4	5	6	7	8	9	10
1. TAKS	1.0									
2. GRADE Passage Comp	.64	1.0								
3. WJ III Passage Comp	.61	.60	1.0							
4. TOWRE Words	.40	.37	.42	1.0						
5. ORF CBM-WF	.36	.32	.37	.68	1.0					
6. ORF CBM-PF	.50	.51	.50	.73	.73	1.0				
7. AIMSweb Maze CBM	.40	.37	.38	.45	.4	.57	1.0			
8. TOSREC	.56	.57	.62	.54	.44	.61	.50	1.0		
9. TOSCRF	.42	.40	.49	.47	.44	.55	.56	.51	1.0	
10. KBIT–2 Verbal Knowledge	.54	.53	.60	.31	.26	.38	.34	.56	.38	1.0
<i>M</i>	2,157	95	93.1	97.1	84.9	130.3	175.0	91.6	92.9	101.2
<i>SD</i>	176.7	11.4	10.5	11.0	24.7	31.6	58.8	13.5	10.7	16.5

Note. $n = 1,421$. TAKS = Texas Assessment of Knowledge and Skills Scale Score; GRADE Passage Comp = Group Reading Assessment and Diagnostic Evaluation Passage Comprehension Subtest Prorated Standard Score; WJ III Passage Comp = Woodcock Johnson III Passage Comprehension Subtest Standard Score; TOWRE Words = Test of Word Reading Efficiency Sight Word Efficiency Standard Score; ORF CBM-WF = Oral Reading Fluency Curriculum Based Measure—Word Fluency Equated Average; ORF CBM-PF = Oral Reading Fluency Curriculum Based Measure—Passage Fluency Equated Average; AIMSweb Maze CBM = AIMSweb Maze CBM Reading Comprehension Total Number Targets; TOSREC = Test of Silent Reading Efficiency and Comprehension; TOSCRF = Test of Silent Contextual Reading Fluency Standard Score; KBIT–2 Verbal Knowledge = Kaufman Brief Intelligence Test–2 Verbal Knowledge Standard Score.

consistent regardless of how comprehension is measured. An exception is that measures of word-level fluency (i.e., ORF words, TOWRE, TOSCRF) appear to be somewhat more strongly related to the WJ III than to the GRADE Passage Comprehension tests. However, the reliabilities of the two measures are significantly different ($p < .05$). The test-retest correlation of WJ III Passage Comprehension from pretest in the fall of 2006 to posttest in the spring of 2007 for our sample was 0.82, whereas the test-retest correlation for the GRADE Passage Comprehension was 0.70. This would attenuate correlations of the GRADE with the fluency measures. The verbal knowledge measure (i.e., KBIT-2), an assessment of vocabulary and world knowledge, was moderately correlated with all comprehension tests and had a somewhat stronger relation with the WJ III passage comprehension test ($r = .60$) than with the other comprehension measures ($r = .53-.54$). Correlations among the three reading comprehension measures were also in the moderate range ($r = .60-.64$), with the strongest relation observed between the GRADE and TAKS, both of which measure comprehension through multiple-choice questions related to passages of various lengths.

To contrast the strengths of the relations between measures of fluency and comprehension, we conducted Z tests of the difference in correlations among these variables (see Table 2). Fourteen comparisons were made for each reading comprehension measure (i.e., TAKS Reading, GRADE Passage Comprehension, and WJ III Passage Comprehension). Each value was compared to the value of 0.0036 (i.e., 0.05 alpha/14 comparisons) to correct for Type I error inflation due to multiple comparisons.

The findings reveal a pattern in which the ORF CBM Passage Fluency and the TOSREC silent reading measure are more highly correlated with the comprehension measures than most other fluency measures. Specifically, ORF CBM Passage Fluency and the TOSREC correlate equally well with TAKS Reading ($Z = 2.22$) and GRADE Passage Comprehension ($Z = 2.26$), although the correlation with WJ III Passage Comprehension is higher for the TOSREC than for ORF CBM Passage Fluency ($Z = 4.67$). The correlations of the TOSREC with each measure of reading comprehension is also stronger than the correlations of TOWRE Sight Word Efficiency ($Z = 5.56-7.38$), ORF CBM Word Fluency ($Z = 6.80-8.95$), the TOSCRF ($Z = 4.92-5.95$), and the AIMSweb Maze CBM ($Z = 5.56-8.64$) with the comprehension measures.

ORF CBM Passage Fluency correlates more strongly with all three comprehension measures than the AIMSweb Maze CBM ($Z = 3.34-4.63$) and ORF CBM Word Fluency ($Z = 4.28-6.15$). The correlations of ORF CBM Passage Fluency with TAKS Reading and GRADE Passage Comprehension are also stronger than those of the TOSCRF ($Z = 2.70$ for TAKS and 3.70 for GRADE) and TOWRE Sight Word Efficiency ($Z = 3.34$ for TAKS and 4.64 for GRADE) with the same comprehension measures.

TABLE 2
Z Test of the Difference in Correlations for Measures of Reading Fluency Predicting
Measures of Reading Comprehension

<i>Comparison</i>	<i>TAKS Reading Comprehension</i>	<i>GRADE Passage Comprehension</i>	<i>WJ III Passage Comprehension</i>
TOSREC vs. ORF CBM-PF	2.22	2.26	4.67 ^a
TOSREC vs. TOWRE	5.56 ^a	6.89 ^a	7.38 ^a
TOSREC vs. ORF CBM-WF	6.80 ^a	8.40 ^a	8.95 ^a
TOSREC vs. TOSCRF	4.92 ^a	5.95 ^a	5.03 ^a
TOSREC vs. AIMSweb Maze	5.56 ^a	6.89 ^a	8.64 ^a
ORF CBM-PF vs. AIMSweb Maze	3.34 ^b	4.63 ^b	3.97 ^b
ORF CBM-PF vs. TOSCRF	2.70 ^b	3.70 ^b	0.35
ORF CBM-PF vs. ORF CBM-WF	4.59 ^b	6.15 ^b	4.28 ^b
ORF CBM-PF vs. TOWRE	3.34 ^b	4.64 ^b	2.70
TOSCRF vs. AIMSweb Maze	0.63	0.94	3.62 ^c
TOSCRF vs. TOWRE	0.64	0.94	2.35 ^c
TOSCRF vs. ORF CBM-WF	1.88	2.45	3.92 ^c
AIMSweb Maze vs. TOWRE	0.00	-0.31	-1.27
AIMSweb Maze vs. ORF CBM-WF	1.24	1.51	0.31

Note. $n = 1,421$. TAKS = Texas Assessment of Knowledge and Skills; GRADE Passage Comp = Group Reading Assessment and Diagnostic Evaluation Passage Comprehension Subtest Prorated Standard Score; WJ III Passage Comprehension = Woodcock Johnson III Passage Comprehension Subtest Standard Score; TOSREC = Test of Silent Reading Efficiency and Comprehension; ORF CBM-PF = Oral Reading Fluency Curriculum Based Measure–Passage Fluency; TOWRE = Test of Word Reading Efficiency Sight Word Efficiency Subtest; ORF CBM-WF = Oral Reading Fluency Curriculum Based Measure–Word Fluency; TOSCRF = Test of Silent Contextual Reading Fluency Standard Score; AIMSweb Maze = AIMSweb Maze Curriculum-Based Measure.

^aCorrelation of TOSREC with measure of reading comprehension is greater than that of the comparison measure. ^bCorrelation of ORF CBM-PF with measure of reading comprehension is greater than that of the comparison measure. ^cCorrelation of TOSCRF with measure of reading comprehension is greater than that of the comparison measure.

Among the other fluency measures, Table 2 shows that the TOSCRF is more strongly correlated with WJ III Passage Comprehension than TOWRE Sight Word Efficiency ($Z = 2.35$), ORF CBM Word Fluency ($Z = 3.92$), and the AIMSweb Maze CBM ($Z = 3.62$).

Accounting for High-Stakes Test Performance

The second question evaluated the extent to which measures of reading fluency and verbal knowledge contribute to the prediction of TAKS reading comprehension performance beyond the contribution of the prior year's performance on the same test. To address this question, linear regression was utilized, and a series of regression models were examined. The approach to analyses was to determine (a) the amount of variance in 2007 TAKS Reading performance that was accounted

for by the 2006 performance on TAKS Reading, (b) whether the addition of the ORF CBM PF (the oral reading fluency measure with the highest correlation with TAKS Reading) accounted for greater variance in 2007 TAKS Reading than 2006 TAKS Reading alone, (c) whether the addition of TOSREC (the silent reading fluency measure with the highest correlation with TAKS Reading) accounted for greater variance in 2007 TAKS Reading than 2006 TAKS Reading alone, (d) whether KBIT-2 Verbal Knowledge accounted for greater variance in 2007 TAKS Reading than 2006 TAKS Reading alone, and (e) whether a model that included all predictors (i.e., 2006 TAKS Reading, ORF CBM PF, TOSREC, KBIT-2 Verbal Knowledge) accounted for more variance than TAKS 2006 Reading alone.

The first model with 2007 TAKS Reading regressed on 2006 TAKS Reading accounted for 44% of the variance $F(1,1301) = 1012.0, p < .0001$. The second model with 2007 TAKS Reading regressed on 2006 TAKS Reading plus ORF CBM PF accounted for 45% of the variance, $F(2,1300) = 539.3$, a statistically significant incremental increase ($p < .0001$). The third model with 2007 TAKS Reading regressed on 2006 TAKS Reading plus TOSREC accounted for 47% of the variance $F(2, 1300) = 580.7$, also a significant increment ($p < .0001$). The fourth model with 2007 TAKS Reading regressed on 2006 TAKS Reading plus KBIT-2 Verbal Knowledge accounted for 47% of the variance, $F(2, 1300) = 582.1$, a significant increment ($p < .0001$). The final model with 2007 TAKS Reading regressed on 2006 TAKS Reading, ORF CBM PF, TOSREC, and KBIT-2 Verbal Knowledge accounted for 49% of the variance, $F(4, 1298) = 317.1, p < .0001$. Thus, additional measures accounted for 1 to 5% of additional unique variance beyond that explained by 2006 TAKS scores in predicting 2007 TAKS scores. Although significant, the amounts of additional variance explained by adding fluency and verbal knowledge measures are small. The model with all predictors accounts for the most variance, with each predictor accounting for significant unique variance.

Classification Accuracy of Oral and Silent Reading Fluency Measures

The third research question addressed whether scores on ORF passage fluency measures, the silent reading measures, and the previous year's TAKS performance adequately classify middle school readers who are at risk for failing a state high-stakes reading comprehension test. To evaluate classification accuracy, we calculated sensitivity, specificity, positive predictive rate, negative predictive rate, and area under the Receiver Operating Characteristic (ROC) curve (generated in SAS 9.0 statistical software) for four fluency measures—ORF CBM Passage Reading Fluency, AIMSweb Maze CBM, TOSREC, and TOSCRF, using various fluency score benchmarks as well as spring 2006 TAKS.

Sensitivity refers to the extent to which a measure correctly classifies struggling readers at risk for reading difficulties. It is calculated by dividing the number

of true positives (those correctly identified as at risk) by the sum of true positives and false negatives (i.e., struggling readers incorrectly classified as proficient). A sensitivity of 100% means that the test identifies all impaired readers as such. Specificity refers to the extent to which a measure correctly classifies readers at low risk for reading difficulties. It is calculated by dividing the number of true negatives (those correctly identified as not at risk) by the sum of true negatives and false positives (proficient readers incorrectly classified as at risk). A specificity of 100% means that the test correctly classifies all proficient readers. There is always a trade-off of false positive and false negative errors, so that adjusting a score cut point to raise either sensitivity or specificity usually results in a reduced value for the other.

The positive predictive rate is the proportion of students with a positive test result (i.e., students who do have reading difficulties) who are correctly identified with a reading difficulty, whereas the negative predictive rate is the proportion of students with a negative result who are correctly identified as not having a reading difficulty. Overall classification accuracy represents the proportion of true or correct classifications. It is calculated by adding the total number of correct classifications (true positives plus true negatives) and dividing by both correct and incorrect classifications (the total number of true positives, true negatives, false positives, and false negatives).

Area under the ROC curve (AUC) represents the ability of a test to correctly classify proficient versus struggling readers. Conceptually, AUC values can be interpreted in several ways: (a) the probability that the test will produce a value for a randomly selected individual without the characteristic that is greater than the value for a randomly chosen individual with a characteristic, (b) the average sensitivity for all possible values of specificity, and (c) the average specificity for all possible values of sensitivity (Lasko et al., 2005). An AUC of 0.5 represents chance classification, whereas an AUC of 1.0 represents perfect classification.

Classification accuracy is evaluated in relation to preestablished benchmarks, or score cut points, used to classify students as proficient versus struggling readers. Table 3 shows the extent to which three different benchmarks on each of the four fluency measures correctly classify readers as at risk or not at risk for failing the TAKS state reading comprehension test. The three oral reading fluency rates evaluated are 100, 130, and 160 words read correctly per minute, which approximate the 25th, 50th, and 75th percentiles in ORF for middle school students according to norms developed by Hasbrouck and Tindal (2006). The Hasbrouck and Tindal ORF norms were utilized because they were created from a large pool of scores (5,546 to 10,520 students in Grades 6–8) from schools in 23 states. The same three percentile ranks (i.e., 25%, 50%, and 75%) were used as benchmarks for the silent reading measures. Score values for each percentile on these measures were calculated according to local norms. For the spring 2006

TABLE 3
 Classification Accuracy Statistics for Reading Fluency Measures Predicting Reader Classification on the Texas Assessment
 of Knowledge and Skills Reading Comprehension Test at Grades 6 to 8^{a,b}

Measure and Benchmark	False-Positive Rate	False-Negative Rate	Sensitivity	Specificity	Positive Predictive Rate	Negative Predictive Rate	Overall Classification Accuracy
25th Percentile ORF	0.07	0.73	0.27	0.93	0.83	0.52	0.58
50th Percentile ORF	0.33	0.34	0.66	0.67	0.70	0.63	0.66
75th Percentile ORF	0.71	0.09	0.91	0.29	0.60	0.74	0.63
25th Percentile Maze	0.14	0.65	0.35	0.86	0.75	0.53	0.59
50th Percentile Maze	0.34	0.37	0.63	0.66	0.68	0.60	0.64
75th Percentile Maze	0.63	0.16	0.84	0.37	0.61	0.67	0.62
25th Percentile TOSREC	0.10	0.63	0.37	0.90	0.81	0.55	0.62
50th Percentile TOSREC	0.34	0.29	0.71	0.66	0.71	0.67	0.69
75th Percentile TOSREC	0.61	0.08	0.92	0.39	0.64	0.80	0.67
25th Percentile TOSCRF	0.13	0.65	0.35	0.87	0.76	0.54	0.59
50th Percentile TOSCRF	0.36	0.37	0.63	0.64	0.67	0.60	0.63
75th Percentile TOSCRF	0.85	0.33	0.67	0.15	0.40	0.34	0.39
TAKS 2006 Reading ^a	0.38	0.18	0.82	0.62	0.55	0.86	0.69

Note. $n = 1,421$. ORF = Oral Reading Fluency Curriculum-Based Passage Fluency; Maze = AIMSweb Maze Curriculum-Based Measure; TOSREC = Test of Silent Reading Efficiency and Comprehension; TOSCRF = Test of Silent Contextual Reading Fluency.

^aBased on 2,150 Scale Score. ^bBase rate = 0.55.

TAKS Reading predictor, we used the benchmark of a scale score of 2,150, representing one half of one standard error of measurement above the state's pass-fail cut point on the test. This benchmark was selected because it includes students whose performance is on the "bubble" between passing and failing the test (i.e., within the test's margin of error). The reading comprehension outcome predicted was spring 2007 TAKS Reading, with performance greater than a scale score of 2,150 defined as "proficient."

Table 3 indicates that, for ORF passage reading fluency, the 50th percentile correctly identifies 66% (i.e., sensitivity) of all students with a reading difficulty and 67% (i.e., specificity) of those who do not have a reading difficulty as measured by the TAKS. At the 50th percentile, the TOSREC has better sensitivity (.71) than ORF passage fluency, whereas the sensitivity of the other two silent fluency measures (.63) is comparable to that of ORF passage reading fluency. The four fluency measures have comparable specificity (range = .64-.67).

Using the 50th percentile on ORF passage reading fluency as the cutoff for prediction of reading risk results in a false positive rate of 33% and a false negative rate of 34%. For the TOSREC, at the 50th percentile, 34% of proficient readers are incorrectly identified as struggling, and 29% of struggling readers are not identified as such.

Lowering the benchmark to the 25th percentile decreases false positive errors but results in false negative errors rates ranging from .63 to .73, meaning that 63% to 73% of struggling readers would not be identified if this benchmark was adopted. Conversely, raising the score cut point on ORF passage reading fluency to the 75th percentile increases the false positive error rate to 71% but substantially reduces the false negative error rate to .09, meaning that nearly all students in need of reading intervention would be identified. A similar pattern is evident for the TOSREC, where using the 75th percentile cutoff results in a high false-positive error rate (61%) and a low false-negative error rate (.08). If fluency measures are used as screening tools, low false-negative rates may be desirable so that all students who need intervention may receive it. The most accurate predictor of 2007 TAKS performance is 2006 TAKS performance. Applying the score cutoff of a 2,150 scale score results in a false-positive error rate of 38% and a fairly low false-negative error rate of 18%.

The AUCs indicate that the TOSREC has the best classification accuracy of any silent reading fluency measure we evaluated, a rate similar to that of individually administered ORF assessments. The AUC for the TOSREC predicting TAKS reading comprehension is 0.76, and the AUC for ORF passage reading fluency for predicting the same measure is 0.73. For both the AIMSweb Maze and TOSCRF, this value is 0.69, a value that is significantly lower than the AUC for the TOSREC. The AUC of 0.82 confirms that previous performance on the TAKS is a strong predictor of future performance on the same measure.

The analysis of AUCs suggests that when predicting pass–fail on the TAKS Reading test, the probability that each ORF or silent reading fluency test will produce a value for a randomly selected student with a reading difficulty that is lower than the value for a randomly selected student without a reading difficulty is significantly better than chance and ranges from 0.69 to 0.76. Previous TAKS performance is, however, the most accurate predictor, with a probability of 0.82 that the student with a reading difficulty will have a lower score than the student without a reading difficulty.

DISCUSSION

Oral Reading Fluency and Reading Comprehension

This study has two important findings with respect to ORF and reading comprehension in Grades 6 to 8. First, like researchers who have examined ORF with elementary-aged students, we found positive and significant relations among measures of ORF and reading comprehension for students in Grades 6 to 8. However, we found that the relations of ORF passage reading and measures of reading comprehension for middle school students are moderate ($r = .50-.51$) and generally weaker than often reported for younger students. Our findings are similar to those reported by Silbergitt et al. (2006) and Torgesen et al. (2003), who found correlations between ORF and comprehension at Grades 6 to 8 in the .50 to .60 range. Compared to the relations of ORF and the Woodcock Reading Mastery Test Passage Comprehension subtest reported at Grades 1 to 4 by Hosp and Fuchs (2005; $r = .79-.84$), our findings for students in middle school indicate appreciably weaker relationships. Thus, this study indicates that the relation between ORF and comprehension differs in older and younger readers.

Second, for middle school students, ORF measured in connected text is more closely related to reading comprehension than ORF measured in word lists, as has been found for younger students. The relation between word list reading fluency and comprehension was generally weak in this study, particularly when comprehension was measured using the GRADE, a standardized norm-referenced test in which students read extended text passages and answer questions about them. Word list fluency was somewhat more strongly related to the WJ III Passage Comprehension subtest, which has been found to be more affected by word-level decoding skills (Francis, Fletcher, Catts, & Tomblin, 2005; Keenan, Betjemann, & Olson, 2008). However, even this relation was significantly weaker than between oral passage reading fluency and comprehension. Thus, for middle school students, oral passage reading fluency may be a reflection of efficiency of text processing both at and beyond the word level, as suggested by Klauda and Guthrie (2008) for fifth-grade students.

Silent Reading Fluency and Reading Comprehension

We found that a group-administered silent reading sentence verification test (TOSREC) is more strongly related to reading comprehension in Grades 6 to 8 than any other silent fluency measure we evaluated; TOSREC also has a stronger relationship with one comprehension measure (i.e., WJ III Passage Comprehension) than individually administered ORF measures. Even so, correlations of the TOSREC and the three comprehension outcomes were moderate ($r = .56-.62$) and similar to those reported by Torgesen et al. (2003), who found relations between the TOSREC and Florida state reading test of .58 at both Grades 6 and 8. These coefficients are substantially lower than those reported by Klauda and Guthrie (2008) at Grade 5 for the WJ III Reading Fluency subtest, which is also a sentence verification task ($r = .72$).

Like Ardoin et al. (2004) and Jenkins and Jewell (1993), who reported that ORF was more strongly related to comprehension than maze tasks in the elementary grades, we found that passage reading ORF was a significantly better predictor of middle school students' comprehension scores than the AIMSweb Maze. In general, the relations between the maze task and reading comprehension measures in this study ($r = .37-.40$) were weaker than those reported by other researchers. For example, Silbergitt et al. (2006) reported correlations in the .50 range at Grades 7 and 8 between maze and comprehension measured through a state high-stakes reading test. Torgesen et al. (2003) found slightly higher correlations between maze and the Florida state reading test at Grades 8 ($r = .59-.63$, depending on how the maze tasks were constructed).

Torgesen et al. (2003) also reported that the TOSCRF was the measure most weakly related to comprehension outcomes at sixth grade (.39) and eighth grade (.22). In our study, the TOSCRF correlated higher than the maze and resulted in correlations of .41 to .50; however, it did not correlate as highly as the TOSREC.

Based on our findings, it appears that sentence verification tasks like the TOSREC, even when administered in groups, hold promise as viable alternatives to individually administered ORF measures for predicting comprehension outcomes at Grades 6 to 8. More research is needed to confirm these findings.

Multiple Measures of Reading Comprehension

We administered three measures of reading comprehension, one in a cloze format (i.e., WJ III Passage Comprehension) and two that require students to answer multiple-choice questions about passages of varying length (i.e., TAKS reading, GRADE Passage Comprehension). There is a tendency in the literature to refer to a single construct called "reading comprehension" that can be measured in various ways (Fletcher, 2006). In our study, correlations among these three measures were

only in the moderate range ($r = .60-.64$), suggesting that the three assessments may be measuring somewhat different domains.

Francis, Fletcher, Stuebing, et al. (2005) and Keenan et al. (2008) reported that reading comprehension measures that utilize a cloze format (e.g., WJ III Passage Comprehension) appear to be more highly correlated with word identification measures than comprehension tests that require students to read more extended text. Our study does not support or contradict these findings. We observed only minimal differences in correlations between measures of word identification fluency and the three measures of reading comprehension, and these results were mixed. TOWRE Sight Word Identification and the ORF word reading CBM were only slightly better predictors of WJ III Passage Comprehension than they were of GRADE Passage Comprehension. Conversely, the correlations between word reading fluency and TAKS reading were slightly higher than the correlations between word reading fluency and the WJ III comprehension measure. Our findings may differ from those previously reported because our measures of word identification were fluency based.

Verbal Knowledge, Fluency, and Reading Comprehension

It has been suggested that vocabulary measures may be useful predictors of reading comprehension at middle school (e.g., Yovanoff et al., 2005). Our findings related to this question were mixed and depended on how reading comprehension was measured. We found moderate relations between a measure of general vocabulary and world knowledge (i.e., KBIT-2 Verbal Knowledge) and reading comprehension. When comprehension was measured with the WJ III Passage Comprehension subtest, the KBIT-2 measure was a better predictor of comprehension ($r = .60$) than was ORF passage fluency ($r = .50$); however, relations between verbal knowledge and the TAKS and GRADE were similar to those between ORF passage fluency and the same comprehension measures. Had we measured vocabulary knowledge with words that were related to the subject matter of the reading comprehension passages, as did Espin and Foegen (1996), it is likely that relations with comprehension outcomes would have been stronger.

Accounting for Variance in High-Stakes Test Performance

Our second research question addressed the amount of variance in the Texas reading comprehension accountability test (i.e., TAKS) that was accounted for by the previous year's scores on the same measure and how much additional variance was explained by adding fluency or verbal knowledge measures. Not surprisingly, we found that a substantial amount of variance in 2007 TAKS scores was explained by the same students' 2006 TAKS scores (44%). Entering single fluency or verbal knowledge measures after 2006 TAKS in multiple regression

models accounted for only minimal additional variance (1–3%) beyond 2006 TAKS alone in predicting 2007 TAKS. Even the model that included two fluency measures and the KBIT–2 Verbal Knowledge measure along with 2006 TAKS accounted for only 5% additional variance. Thus, middle school students' performance on the state reading comprehension accountability test in one year is best predicted by the previous year's performance on the same test. Adding a fluency measure, whether that is ORF or a silent measure, does not appear to contribute meaningfully to the accuracy of the prediction.

Identification of Students in Need of Intervention

More relevant in a practical sense is the question of how accurately these measures classify students as proficient readers and students with reading difficulties so that students who need reading intervention may receive it. Compton, Fuchs, Fuchs, Elleman, and Gilbert (2008) observed that the use of screening measures to identify students in the primary grades at risk for reading difficulties requires that these measures “yield a high percentage of true positives . . . while identifying a manageable risk pool by limiting false positives” (p. 329). These goals apply just as well to the identification of older students in need of intervention. In fact, the consequences of misidentification may be even more serious at higher grades, when students may be several years behind their peers and have little time to close that gap. Some suggest cut scores be applied to screening measures that will maintain sensitivity rates of 90% to 95% (e.g., Jenkins, 2003) in order to limit the numbers of struggling readers who are not identified.

Our analyses explored the use of ORF passage reading and three silent reading measures (administered in the fall) and the previous spring administration of the TAKS reading comprehension test to identify students who do not pass the TAKS reading comprehension test administered in the spring of Grades 6, 7, and 8. Based on the analyses of area under the ROC curves, all of the fluency measures identify students in need of reading intervention at a rate higher than chance, and can be classified as “fair” predictors of TAKS performance for middle school readers. However, to maintain sensitivity levels of 90% or higher on these fluency measures, very high cut scores must be adopted. Applying these stringent benchmarks would result in large numbers of adequate readers who “fail” the screen and appear to be struggling readers (i.e., false positive errors). If cut scores on these measures are adjusted to limit false-positive errors, false-negative errors will increase to what may be unacceptable levels, and many students in need of intervention will be “missed.” Because we sampled students at a level within 0.5 standard error of measurement of the passing score of the TAKS (roughly the 30th percentile), the definition of “struggling readers” is relatively liberal. Making the cut point more stringent would also increase false negatives, paralleling the effects of adjusting the cut points in Table 3.

For the fluency measures, the cut score that best limited both false positive and false negative errors was the 50th percentile, where one might reasonably expect that half of the population would meet the benchmark and half would not. Most middle schools would be hard-pressed to provide reading intervention classes to half of their students, and, in most schools, many of these students would not require this kind of intervention. Even at this level, about 30% to 40% of both adequate and inadequate comprehenders would be misidentified. Thus, the use of fluency measures as the sole indicators of potential comprehension difficulties in students in Grades 6 to 8 may be a weak approach.

The measure with the best prediction accuracy for the spring 2007 TAKS administration was the spring 2006 TAKS. The AUC was .82, which exceeded all other measures. In exploratory analyses, we investigated whether prediction accuracy would be increased by combining fluency measures with 2006 TAKS reading, although our multiple regressions predicting 2007 TAKS reading comprehension indicated that entering measures of oral reading fluency, silent reading fluency, and verbal knowledge along with 2006 TAKS Reading explained only 5% of additional variance beyond 2006 TAKS alone. Not surprisingly given these small increments in explained variance, prediction models that included fluency and verbal knowledge in addition to 2006 TAKS reading led to only slight increments in increased classification accuracy. For example, entering all the predictors in a single model led to a 2% improvement in the false-negative rate at the cost of a slight increase in the false-positive rate (2%) and a slight increase in overall accuracy (1%) relative to 2006 TAKS Reading alone. However, there may be diagnostic utility in following the reading comprehension test with a fluency probe.

ORF as a Diagnostic Tool

If the previous year's state test scores are used to screen students, which these results support, all that is known without additional assessment is that students are at risk for failing the subsequent year's high-stakes reading test. This does not indicate the nature of their reading problems, information that is needed if reading intervention teachers are to target students' areas of need in order to make efficient and effective use of limited intervention time. Barth, Denton, Cirino, Fletcher, and Francis (2010) examined whether oral reading fluency measures could assist in diagnosing the types of reading difficulties that prevent successful reading for middle school students. Middle school readers were defined as having comprehension difficulties, fluency difficulties, decoding difficulties, or combinations of these, based on external measures of decoding, fluency, and comprehension. Results suggested that readers with decoding, oral reading fluency, and reading comprehension difficulties performed approximately 2.5 *SDs* below students with only comprehension difficulties on ORF CBM passage fluency.

Results also suggested that readers with primarily reading fluency and comprehension difficulties, but stronger decoding, performed approximately $.5 SD$ below students with only comprehension difficulties on the same measure. Thus, fluency rates help differentiate these three groups of struggling readers who clearly have different instructional needs. For middle school readers who struggle with decoding and read connected text at very slow rates, interventions that target both word analysis skills and reading comprehension are likely the most efficacious. Students with ORF passage reading levels that are somewhat below average, but stronger decoding skills, may benefit from fluency-oriented intervention along with explicit instruction in reading comprehension. For students who read connected text at more grade-appropriate fluency levels, but continue to struggle with reading comprehension, interventions that target verbal knowledge and the integration and use of meaning embedded in text are likely to be more beneficial.

Study Limitations

A limitation of this study is that we only evaluated the prediction accuracy of fluency measures using three rigid ad hoc cut points. It is possible that predictive accuracy would be stronger if we had actively investigated cut points that maximized predictive utility (Fletcher, Schmidt, & Satz, 1979). In addition, establishing confidence intervals that would take into account measurement error around the cut point might also improve predictive accuracy (Francis, Fletcher, Stuebing, et al., 2005). Finally, the generalizability of our findings may be somewhat limited because we over-sampled students with reading difficulties; however, no distributions we examined differed significantly from the normal distribution.

Implications for Practice and Future Research

Because of the limited research conducted with secondary-level students with reading difficulties, there has been a tendency to generalize findings with younger children to older students. Our study demonstrates that such generalization may not always be appropriate. There is a need for research that investigates assessment and intervention practices directly with middle and high school students.

Implications for assessment. Based on our findings, educators of middle school students may be best able to identify students at risk for failing state high-stakes reading comprehension tests by examining students' performance on these tests during the previous school year. The use of these assessments would need to be evaluated in each different state, but the advantage in terms of reducing the need for additional assessment of all students in middle schools

is significant. Because all that would be known from this approach is that the student did not pass a comprehension test, following this initial screening with fluency assessments may enable teachers to quickly diagnose the nature of students' reading problems. Poor comprehension in older students can be related to complex comprehension of difficulties in word reading, reading fluency, oral language comprehension, and construction of meaning from text. This makes reading intervention with older students challenging, and teacher' use of fluency measures may help them provide these students with effective instruction. More research directly addressing the use of both oral and silent fluency measures for this purpose is needed.

This study did not investigate the use of fluency measures to monitor reading progress in middle school students. Repeated assessment using brief measures of fluency may be useful for progress monitoring. However, there is a need for research investigating whether such measures are valid for this purpose for all middle school students with reading difficulties, or primarily for those who have word reading and fluency difficulties.

There is also a need for research that examines how progress in reading comprehension might be validly and reliably monitored for students in the secondary grades. In addition, practitioners and researchers would benefit from the development and validation of reliable assessments of reading comprehension that could be used to screen students to identify those with reading comprehension difficulties, as well as diagnostic measures that could be used to identify with greater specificity the dimensions of reading comprehension that should be targeted in intervention (e.g., recalling text details, making different kinds of inferences). Development of such measures would require a deeper understanding of the construct of reading comprehension and how various aspects of comprehension might be measured.

Implications for instruction. Besides addressing the identification of middle school students in need of reading intervention, this study has implications for their instruction. With young children, theoretical models such as that developed by LaBerge and Samuels (1974) suggest that providing fluency instruction will impact reading comprehension as children build automaticity with decoding processes. Our study, and others, illustrate that this assumption is not necessarily valid at Grades 6–8, when the relation between fluency and comprehension is weaker than often observed with young children. Older students with reading comprehension difficulties are likely to need interventions that directly address vocabulary, world knowledge, and comprehension processes. Rather than addressing only decoding and fluency, it is likely that interventions for secondary-level students who struggle with reading comprehension must include extensive, explicit instruction in making meaning from text. The field would benefit from research focused on the effectiveness of such interventions for secondary

school students who have impaired comprehension with or without accompanying difficulties in decoding and fluency.

ACKNOWLEDGMENT

This research was supported by grant P50 HD052117 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health.

REFERENCES

- Ardoin, S. P., Witt, J. C., & Suldo, S. M. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33*, 218–233.
- Barth, A. E., Denton, C. A., Cirino, P. T., Fletcher, J. M., & Francis, D. (2010). *The diagnostic utility of oral reading fluency measures for middle school students*. Manuscript in preparation.
- Biancarosa, G., & Snow, C.E. (2004). *Reading next—A vision for action and research in middle and high school literacy: A report from Carnegie of New York*. Washington, DC: Alliance for Excellence in Education.
- Compton, L., Fuchs, D., Fuchs, L. S., Elleman, A. M., & Gilbert, J. K. (2008). Tracking children who fly below the radar: Latent transition modeling of students with late-emerging reading disability. *Learning and Individual Differences, 18*, 329–337.
- Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children, 62*, 497–514.
- Fewster, S., & Macmillan, P. D. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education, 23*, 149–156.
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*, 323–330.
- Fletcher, J. M., Schmidt, R. K., & Satz, P. (1979). Discriminant function strategies for the kindergarten prediction of reading disabilities. *Journal of Clinical Neuropsychology, 1*, 151–166.
- Francis, D. J., Fletcher, J. M., Catts, H. W., & Tomblin, T. J. (2005). Dimensions affecting the assessment of reading comprehension. In S. A. Stahl & S. G. Paris (Eds.), *Children's reading comprehension and assessment* (pp. 369–394). Mahwah, NJ: Erlbaum.
- Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of learning disabilities: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38*, 98–108.
- Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology, 46*, 315–342.
- Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 61*, 15–24.

- Fuchs, L. S., Fuchs, D., Hosp, M. K., & 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.
- Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, 9, 20–28.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.
- Hale, A. D., Skinner, C. H., Williams, J., Hawkins, R., Neddenriep, C. E., & Dizer, J. (2007). Comparing comprehension following silent and aloud reading across elementary and secondary students: Implication for curriculum-based measurement. *Behavior Analyst Today*, 8(1), 9–23.
- Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2006). *Test of Silent Contextual Reading Fluency*. Austin, TX: Pro-Ed.
- Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *Reading Teacher*, 59, 636–644.
- 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, 9–26.
- Howe, K. B., & Shinn, M. M. (2002). *Standard reading assessment passages (RAPs) for use in general outcome measurement: A manual describing development and technical features*. Eden Prairie, MN: Edformation.
- Jenkins, J. R. (2003). *Candidate measures for screening at-risk students*. Paper presented at the Conference on Response to Treatment as Learning Disabilities Identification, Kansas City, MO. Retrieved from <http://www.nrcld.org/symposium2003/jenkins/index.html>
- Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading Aloud and Maze. *Exceptional Children*, 59, 421–432.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test* (2nd ed.). Minneapolis, MN: Pearson Assessment.
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12, 281–300.
- Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, 100, 310–321.
- Lasko, T. A., Bhagwat, J. G., Zou, K. H., & Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, 38, 404–415.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.
- Lennon, C., & Burdick, H. (2004). *The Lexile Framework as an approach for reading measurement and success*. MetaMetrics, Inc.
- Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18–78). New York, NY: Guilford.
- Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2004). *Test of Silent Word Reading Fluency*. Austin, TX: Pro-Ed.
- McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, 33, 193–203.
- Miller, S. D., & Smith, D. E. (1990). Relations among oral reading, silent reading, and listening comprehension of students at differing competency levels. *Reading Research and Instruction*, 29, 73–84.
- Paris, S. G., Carpenter, R. D., Paris, A. H., & Hamilton, E. E. (2005). Spurious and genuine correlates of children's reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 131–160). Mahwah, NJ: Erlbaum.

- Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42*, 546–567.
- Roehrig, H. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46*, 343–366.
- Schatschneider, C., Buck, J., Torgesen, J., Wagner, R., Hassler, L., Hecht, S., & Powell-Smith, K. (2004). *A multivariate study of individual differences in performance on the reading portion of the Florida Comprehensive Assessment Test: A brief report*. Retrieved from Florida Center for Reading Research Web site. http://www.fcrr.org/technicalreports/multi_variate_study_december2004.pdf
- Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zheng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal, 107*, 429–448.
- Shinn, M., Knutson, N., Collins, V. L., Good, R., & Tilly, W. D. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459–479.
- Shinn, M. R., & Shinn, M. M. (2002). *AIMSweb training workbook*. Eden Prairie, MN: Edformation.
- Silbergliitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43*, 527–535.
- Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on state mandated performance based assessment using oral reading fluency. *School Psychology Review, 30*, 407–419.
- Texas Education Agency (2004a). *Appendix 20—Student Assessment Division Technical Digest 2004–2005*. Retrieved from <http://ritter.tea.state.tx.us/student.assessment/resources/techdig05/appendices.html>
- Texas Education Agency. (2004b). *Student Assessment Division Technical Digest 2004–2005*. Retrieved from <http://ritter.tea.state.tx.us/student.assessment/resources/techdig05/index.html>
- Torgesen, J., Nettles, S., Howard, P., & Winterbottom, R. (2003). *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* (FCRR Report No.6). Tallahassee, FL: Florida Center for Reading Research at Florida State University. Retrieved from http://www.fcrr.org/TechnicalReports/Progress_monitoring_report.pdf
- Torgesen, J., Wagner, R., & Rashotte, C. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.
- University of Houston. (2008). *Texas Middle School Fluency Assessment*. Houston, TX: Author.
- Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). *The relationship between oral reading fluency and Ohio proficiency testing in reading*. Eugene, OR: University of Oregon.
- Vaughn, S. R., Cirino, P. T., Wanzek, J., Wexler, J., Fletcher, J. M., Denton, C. A., . . . Francis, D. J. (2010). Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review, 39*(1), 3–21.
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *Test of Silent Reading Efficiency and Comprehension*. Austin, TX: Pro-Ed.
- Wayman, M. 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.
- Williams, K. T. (2001). *Group Reading Assessment Diagnostic Evaluation*. Shoreview, MN: Pearson AGS Globe.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Achievement*. Itasca, IL: Riverside.
- Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*, 4–12.

