Reading skill components and impairments in middle school struggling readers

Paul T. Cirino,
Department of Psychology, University of Houston, 4800 Calhoun St., Houston, TX 77204-5053, USA; Texas Institute for Measurement, Evaluation, and Statistics (TIMES), Houston, TX, USA

Melissa A. Romain,
Texas Institute for Measurement, Evaluation, and Statistics (TIMES), Houston, TX, USA

Amy E. Barth,
Texas Institute for Measurement, Evaluation, and Statistics (TIMES), Houston, TX, USA

Tammy D. Tolar,
Texas Institute for Measurement, Evaluation, and Statistics (TIMES), Houston, TX, USA

Jack M. Fletcher, and
Department of Psychology, University of Houston, 4800 Calhoun St., Houston, TX 77204-5053, USA; Texas Institute for Measurement, Evaluation, and Statistics (TIMES), Houston, TX, USA

Sharon Vaughn
The University of Texas at Austin, 1 University Station, D4900, Austin, TX 78712-0365, USA

Abstract

This study investigated how measures of decoding, fluency, and comprehension in middle school students overlap with one another, whether the pattern of overlap differs between struggling and typical readers, and the relative frequency of different types of reading difficulties. The 1,748 sixth, seventh, and eighth grade students were oversampled for struggling readers (n = 1,025) on the basis of the state reading comprehension proficiency measure. Multigroup confirmatory factor analyses showed partial invariance among struggling and typical readers (with differential loadings for fluency and for comprehension), and strict invariance for decoding and a combined fluency/comprehension factor. Among these struggling readers, most (85 %) also had weaknesses on nationally standardized measures, particularly in comprehension; however, most of these also had difficulties in decoding or fluency. These results show that the number of students with a specific comprehension problem is lower than recent consensus reports estimate and that the relation of different reading components varies according to struggling versus proficient readers.

Keywords

Introduction

Adolescent literacy has emerged as a major problem for research and instruction over the past decade, with approximately six million adolescents recognized as reading below grade

level (Joftus & Maddox-Dolan, 2003; Vaughn, Denton, & Fletcher, 2010b). National, state, and local reports reveal that adolescent struggling readers score in the lowest percentiles on reading assessments (Grigg, Daane, Jin, & Campbell, 2003; Deshler, Schumaker, Alley, Warner, & Clark, 1982; National Assessment of Educational Progress [NAEP], 2009) and show significant deficits in word reading accuracy, oral reading fluency, and reading comprehension (Fuchs, Fuchs, Mathes, & Lipsey, 2000; Hock et al., 2009; Vellutino, Fletcher, Snowling, & Scanlon, 2004; Vellutino, Tunmer, Jaccard, & Chen, 2007).

Fueled in part by the significant growth in research and the translation of research into instruction for beginning reading, and the growing number of adolescents reading 4–6 years below grade level, increased attention is being paid to students beyond the early grades who continue to struggle with reading. Children in the early grades struggle with basic reading processes involving decoding, but over time these skills should be mastered as proficiency develops and the focus shifts to comprehension. The Reading Next report (Biancorosa & Snow, 2006) underscored this transition in reading proficiencies, and the changing needs of struggling readers, by suggesting that "only 10 percent of students struggle with decoding (reading words accurately)" (p. 11), and that "Some 70 % of older readers require some form of remediation. Very few of these older struggling readers need help to read the words on a page; their most common problem is that they are not able to comprehend what they read" (p. 3). However, empirical data on the patterns of reading difficulty in older struggling readers is limited.

Components of reading skills

Assertions about the incidence of comprehension impairments in secondary level struggling readers presume that reading can be separated into specific components across multiple grade levels and that these components have different developmental trajectories. The well-known simple view of reading (Gough & Tunmer,1986) has provided a framework for numerous studies in which skill difficulties in poor readers have been investigated for early elementary (Adlof, Catts, & Lee, 2010; Catts, Hogan, & Fey, 2003; Leach, Scarborough, & Rescorla, 2003; Nation, Clarke, Marshall, & Durand, 2004; Shankweiler et al., 1999; Torppa et al., 2007) and middle school students (Adlof et al., 2010; Buly & Valencia, 2002; Catts, Adlof, & Weismer, 2006; Catts, Hogan, & Adlof, 2005; Hock et al., 2009; Leach et al., 2003). The simple view proposed that reading comprehension was essentially the product of decoding and listening comprehension components. As decoding develops, reading comprehension becomes more aligned with listening comprehension, but either component represents a potential constraint interfering with reading comprehension. The hypothesized components of the simple view have been supported through many studies (Kirby & Savage, 2008). For example, regression and structural equation modeling investigations report that most variance in reading comprehension can be accounted for by word decoding and listening comprehension (Aaron, Joshi, & Williams, 1999; Catts et al., 2005; Cutting & Scarborough, 2006; Kendeou, van de Broek, White, & Lynch, 2009; Sabatini, Sawaki, Shore, & Scarborough, 2010).

In addition to predicting variance in reading comprehension, a distinction highlighted by the simple view is that it is not always the case that students with reading difficulties struggle with reading comprehension and word decoding concurrently—some students may have word level difficulties but have high levels of comprehension, while other students may have good word level skills yet have reading comprehension difficulties (otherwise termed as specific reading comprehension difficulties; Catts et al., 2006; Share & Leikin, 2004; Torppa et al., 2007). Intervention studies have also demonstrated that although improvements in word decoding difficulties can be made, the effects do not always transfer to reading comprehension for struggling readers (Calhoon, Sandow, & Hunter, 2010).

The National Reading Panel [NRP] (2000) identified five targets for instruction to enhance proficiency in reading: phonemic awareness, phonics (decoding), comprehension, fluency, and vocabulary. Phonologicalawareness and phonics are clearly tied to the development of word recognition skills, whereas vocabulary and comprehension are tied together as the comprehension component. Fluency is the fifth component, although its role requires further investigation. For example, fluency is likely to be an outgrowth and extension of word recognition, but may also represent the speed by which the reader is effectively able to generate a meaningful representation of text (Perfetti, et al., 2008; van den Broek et al., 2005). It is clear, however, that students who are accurate but slow in decoding can be identified (Wolf & Bowers, 1999), and that different types of fluency may play different roles for comprehension (Kim, Wagner, & Foster, 2011). Few studies consider fluency as a potentially separable component of reading ability despite the evidence that in languages with more transparent relations of orthography and phonology, problems with decoding accuracy may be less likely to characterize students with reading problems at the level of the single word than fluency of reading words and text, and spelling (Ziegler & Goswami, 2005).

Much of the research on decoding, fluency, and comprehension components of reading has focused on the early elementary grades. While this has no doubt led to earlier identification of struggling students, many students with reading disabilities are not identified until their late elementary and adolescent years (Hock et al., 2009; Leach et al., 2003). Identifying specific skill deficits in reading (word reading, fluency, comprehension) and in reading-related processes (e.g., vocabulary, listening comprehension) is one way to begin to address reading difficulties more broadly construed in older struggling readers. As one step toward this goal, the present study is specifically focused on combinations of direct reading difficulties among middle grade readers.

Patterns of skill impairments

The literature regarding reading development in readers prior to middle school supports the differentiation of skills among struggling readers. For example, Aaron et al. (1999) identified four subgroups of poor readers in Grades 3, 4, and 6: only decoding, only comprehension, both decoding and comprehension, and orthographic processing/speed (the latter possibly a fluency-impaired group at the oldest age). Several studies have shown that children more typically exhibit difficulties in word recognition and that only a small percentage (about 6–15 % of poor readers) exhibit difficulties specific to reading comprehension (Catts et al., 2003; Leach et al., 2003; Nation et al., 2004; Shankweiler et al., 1999), consistent with a 15 % estimate put forth by Nation (1999). Such students are, however, identifiable, particularly as children develop into more fluent readers, and reading comprehension emerges as a distinct, yet still closely related, skill from word recognition (Nation, 2005; Storch & Whitehurst, 2002). Leach et al. (2003) also found that most of their participants with specific comprehension problems were identified later (grades 4–5) in school, but even in that Grade 4–5 study, the number of children with problems specific to comprehension was under 20 %.

Rates of specific difficulties in decoding are much more variable. Torppa et al. (2007) followed a sample of 1,750 Finnish first- and second-graders. Of students identified as having reading difficulties, 51 % were slow decoders, 22 % were poor comprehenders, and the remainder were weak in both. Leach et al. (2003) and Catts et al. (2003) found that 42 and 36 % of their poor readers to have specifically weak decoding, whereas Shankweiler et al. (1999) found only 18 % of their poor readers to be specifically weak in decoding.

Although specific incidence rates vary with sample characteristics, and according to which reading skills are assessed, it is clear that decoding and comprehension difficulties occur

together at all ages, and are correlated, but separable dimensions of reading ability (Snowling & Hulme, 2012). By middle school though, decoding skills are expected to be well-developed, allowing for greater differentiation of reading components in the areas of fluency and comprehension. Although there are few studies of component impairments in middle or high school, Hock et al. (2009) evaluated reading components in a sample of 345 eighth- and ninth-grade urban students. The sample was selected to represent similar numbers of readers across the range of proficiency (unsatisfactory, basic, proficient, advanced, and exemplary) on their state (Kansas) yearly progress test, which assesses comprehension, fluency, and decoding, as well as other facets. Principal component analyses of their reading related measures yielded composites of word level, fluency, vocabulary, and comprehension skills. Struggling readers included those falling below the 40th percentile. There were 202 struggling readers; of these students, 61 % of students exhibited difficulty in all four of these component areas, whereas less than 5 % had difficulties only in comprehension (and 10 % with difficulties in comprehension and/or vocabulary).

Limitations of previous research

In addition to the need for more studies regarding reading components in older students, the current literature: (a) has a focus on observed (vs. latent) measures of reading, (b) lacks systematic comparisons of components in struggling versus typical readers, and (c) has a limited focus on fluency relative to decoding and comprehension skills.

Latent versus observed variables—Cutting and Scarborough (2006) have demonstrated how task parameters influence the different patterns of relations among observed measures depending on the measure of reading comprehension utilized (Gates-MacGinitie Reading Test—Revised [G-M; MacGinitie, MacGinitie, Maria, & Dreyer, 2000]; Gray Oral Reading Test—Third Edition [GORT-3; Wiederhold & Bryant, 1992]; Wechsler Individual Achievement Text [WIAT; Psychological Corporation, 1992]). For example, decoding skills accounted for twice as much variance in students' WIAT performances (11.9 %) relative to the G-M or GORT-3; conversely, oral language skill accounted for more variance in the G-M (15 %) relative to either WIAT or GORT-3 performances (9 %). Furthermore, Rimrodt, Lightman, Roberts, Denckla, and Cutting (2005) found that 25 % of a subgroup of Cutting and Scarborough's (2006) sample were identified with comprehension difficulties by all three reading comprehension measures and only about half of the subgroup were identified by any single one of the measures. Therefore, any given assessment may yield different groups of readers, making it difficult to ascertain the "best" way to assess reading comprehension.

Despite the difficulties in using individual measures, few studies have approached reading components from a latent variable or a combined measurement perspective, with fewer still in older students. For example, Pazzaglia etal. (1993) reported on a series of studies (including factor analytic work) that also distinguished between decoding and comprehension, with their decoding factor including measures of both reading accuracy and naming speed. However, these studies focused on younger students. Torppa et al. (2007) used factor mixture modeling to derive latent classes of students based on measures of fluency and comprehension in their large sample, but did not examine multiple measures at each time point, or separate fluency from decoding, and again focused on younger readers. Kendeou et al. (2009a, b) also identified decoding and comprehension factors, although the former included phonological awareness and vocabulary, and the latter was assessed auditorily, given that the children were aged four to six. In the second study reported in Kendeou et al., two similar factors were derived for Grade 1 students, with vocabulary, nonword fluency, and connected text identified as decoding, with retell fluency cross-loading on both decoding and comprehension factors. Nation and Snowling (1997) found

evidence for a two-factor model, also in younger children (ages seven to 10) corresponding to decoding (measures of word reading accuracy with context, word reading without context, and non-word reading) and comprehension (measures of narrative listening and text comprehension). Francis et al. (2005) differentiated among two measures of reading comprehension, the Woodcock Johnson-R Passage Comprehension (WJ PC) and the Diagnostic Assessment of Reading Comprehension (DARC) in their sample of 3rd grade English language learners. In a latent variable framework, these reading comprehension measures showed differential prediction from factors of decoding, fluency, oral language, phonological awareness, memory, and nonverbal IQ; a primary difference was the stronger relation of decoding and fluency to WJ PC relative to the DARC.

There have been some investigations that utilized a latent variable approach with older readers. Vellutino et al. (2007) utilized multi-group (2nd–3rd graders, 6th–7th graders) confirmatory factor analyses with 14 measures. Cognitive factors of phonological coding, visual coding, phonologically-based skills, knowledge, visual analysis, and spelling, had direct and indirect influences on decoding and language comprehension, which in turn exerted effects on reading comprehension. Decoding was a significant path in younger but not older readers, whereas language comprehension, while significant in both groups, was stronger in the older group. On the other hand, Sabatini et al. (2010), in a confirmatory factor analysis with a large sample of adults with low literacy, found that factors of decoding and listening comprehension were sufficient to account for reading comprehension (although their fluency and vocabulary factors were separable but highly related to decoding and listening comprehension, respectively). Buly and Valencia (2002) examined 108 4th grade students who did not pass a state test. Using cluster analyses, they found 17.6 % (clusters 1 and 2) to be low in only comprehension, 24.1 % (clusters 5 and 6) to be low in only fluency, 17.6 % (cluster 4) were low in decoding and fluency, 16.7 % (clusters 7 and 8) were low in fluency and comprehension, 14.8 % (cluster 3), were low in decoding and comprehension, and 9.3 % (clusters 9 and 10) were low in all three areas. Brasseur-Hock, Hock, Kieffer, Biancarosa, and Deshler (2011), with the sample of Hock et al. (2009), used latent class analyses to subgroup their sample of 195 below-average comprehenders; 50 % had global weaknesses in reading and language (severe or moderate), 30 % were specifically weak in fluency, and approximately 10 % each in language comprehension and reading comprehension. The studies reviewed above vary along a number of dimensions, including age, sample size, and types of measures, but do show both overlap and separation of decoding and comprehension skills to varying degrees.

Struggling versus typical readers—Individual reading skills may also correlate differentially within struggling versus typical readers. On one hand, if skills are more homogeneous in typical readers (meaning that they tend to do well in word recognition, fluency, and reading comprehension), then it is particularly important to differentiate among strugglers. On the other hand, if skills are more homogeneous in strugglers (meaning that the struggling readers tend to do poorly in word recognition, fluency, and reading comprehension), then the approach to remediation should not need to vary much, or perhaps should add a focus on non-reading factors (i.e., motivation/engagement). Thus, evaluating such patterns through techniques such as measurement invariance (e.g., Meredith, 1993) would be of benefit.

Fluency as a component—Fluency is clearly an essential reading component, given its identified role in the NRP report (2000), and its frequent use in progress monitoring (Espin & Deno, 1995; Graney and Shinn, 2005; Hasbrouck & Ihnot, 2007; Shinn, Good, Knutson, Tilley, & Collins, 1992). Aaron et al. (1999) identified a subgroup that could be associated with fluency (their orthographic/processing speed subgroup). Other investigators (Adlof, Catts, & Little, 2006; Hock et al., 2009) found that that fluency-specific problems were rare,

and that many adolescent struggling readers exhibit difficulties in all components, including fluency. Kendeou et al. (2009a, b) found that measures of fluency loaded with both decoding and comprehension factors in their young students. Kim et al. (2011) addressed the role of the fluency component in a latent variable context, and differentiated contributions among struggling and typical readers. They found that oral reading was a stronger predictor of comprehension than silent reading, and also that listening comprehension was more important than decoding fluency for struggling readers, with the opposite pattern in average readers. Kim et al. focused on 1st graders, and therefore did not address the same questions as this study, but it does exemplify the benefit of attending to some of the issues raised above.

The current study

The present study addresses several gaps in the literature by: (a) examining the diversity of reading skills in adolescent readers, and doing so in a latent variable framework, (b) comparing the way these skills are related in struggling versus typical readers, and (c) evaluating the extent to which subtypes of readers exhibit difficulty across word recognition, fluency, and reading comprehension components.

Hypotheses

1. We expect that individual reading measures would serve as indicators of latent constructs of decoding, fluency, and comprehension, according to their traditional status. Competing models will assess the extent to which measures that emphasize all three skills comprise a separate factor, or are best construed as indicators of one of the three aforementioned constructs. Although we clearly expect the reader groups to have different latent means, we expect that the factor structure will be invariant across reader type (struggling vs. typical); in other words, that measures assessing reading components do so in the same manner in both types of readers. 2. Among struggling readers defined by their performance on the Texas Assessment of Knowledge and Skills (TAKS; Texas Educational Agency, 2006a, b, described below) test, which emphasizes comprehension, we expect that the vast majority of these students will have difficulties in one or more external reading measures as gauged by nationally normed measures (<25th percentile); we expect these difficulties to be in multiple areas rather than in comprehension alone; and we expect that a sizeable proportion will continue to have weaknesses with basic decoding skills.

Method

Participants

School sites—This study was conducted in seven middle schools from two urban cities in Texas, with approximately half the sample from each site. Three of the seven schools were from a large urban district in one city with campus populations ranging from 500 to 1,400 students. Four schools were from two medium size districts (school populations ranged in size from 633 to 1,300) that drew both urban students from a nearby city and rural students from the surrounding areas. Based on the state accountability rating system, two of the schools were rated as recognized, four were rated as acceptable, and one school was rated academically unacceptable (though had been rated as acceptable at the initiation of the study). Students qualifying for reduced or free lunch ranged from 56 to 86 % in the first site, and from 40 to 85 % in the second site.

Students—The current study reports on 1,785 sixth through eighth grade students who were assessed in the Fall of the 2006–2007 academic year and who were part of the middle school portion of a larger project on learning disabilities (http://www.texasldcenter.org). The only exclusion criteria were: (a) enrollment in a special education life skills class; (b) took

the reading subtest of the SDAA II at a level lower than 3.0; (c) had a documented significant disability (e.g., blind, deaf); or (d) received reading and/or language arts instruction in a language other than English. As part of the larger study, students were randomized to various treatment conditions, but the data reported here was collected prior to any intervention. In all, of the 1,785 students, 37 were excluded for the above reasons.

The final sample of 1,748 students was composed of 1,025 struggling readers; this represented all students who did not meet criteria on the state reading comprehension proficiency assessments, and 723 randomly selected typical readers who were assessed at the initial Fall time point. "Struggling" readers were defined as students who either (a) scored below the benchmark scale score of 2,100 on their first attempt in Spring of 2006 on the TAKS measure); (b) performed within the upper portion of one standard error of measurement surrounding the TAKS cut-point (i.e., scale scores ranging from 2,100 to 2,150 points); or (c) were in special education and did not take the reading subtest of TAKS, but rather took the reading subtest of the State Developed Alternative Assessment II (SDAA II; Texas Education Agency, 2006b; c). Students designated as "Typical" readers achieved scores above 2150 on TAKS on their first attempt in Spring, 2006. We selected all struggling readers to better generalize to this population, given the purposes of the intervention project from which these students originated. We randomly selected typical readers who passed TAKS to represent approximately two-thirds the number of struggling readers, or 60 % struggling and 40 % typical readers; the final constituted sample was composed of 59 % struggling readers and 41 % typical readers.

Individual school samples ranged from 103 to 544; one site had 957 participants (55 %), and the other had 791. There were 675 (39 %) students in Grade 6, 410 (23 %) in Grade 7, and 663 (38 %) in Grade 8. The proportion of struggling versus typical readers did not differ across grade or site (both $p > .05$). The sample was diverse, as shown in Table 1.

Measures

All examiners completed an extensive training program conducted by the investigators regarding test administration and scoring procedures for each task within the assessment battery. Each examiner demonstrated at least 95 % accuracy at test administration during practice assessments prior to testing study participants. Testing was completed at the student's middle school in quiet locations designated by the school (i.e., library, unused classrooms, theatre). Measures are described according to the domain within which they were used, and are further described at:http://www.texasldcenter.org/research/project2.asp.

Criterion for struggling readers—The Texas Assessment of Knowledge and Skills (TAKS) Reading subtest (Texas Educational Agency, 2006a, b) was the Texas academic accountability test when this study was conducted. It is untimed, criterion-referenced, and aligned with grade-based standards from the Texas Essential Knowledge and Skills (TEKS). In the reading subtest, students read expository and narrative texts of increasing difficulty, and answer multiple choice questions designed to measure students' understanding of the literal meaning of the passages, vocabulary, as well as aspects of critical reasoning. Standard scores are the dependent measure used in this study.

Decoding—Woodcock-Johnson-III Tests of Achievement (WJ-III; Woodcock et al., 2001) Letter-Word Identification and Word Attack subtests. The WJ-III is a nationally standardized individually administered battery of achievement tests with excellent psychometric properties (McGrew & Woodcock, 2001). Letter Word Identification assesses the ability to read real words and Word Attack assesses the ability to read nonwords. Standard scores for these subtests were utilized.

Fluency—The Test of Word Reading Efficiency (TOWRE: Torgesen et al., 2001), with Sight Word Efficiency (SWE) and Phonemic Decoding Efficiency (PDE) subtests, are individually administered tests of speeded reading of individual words presented in list format; subtests are also combined into a composite. The number of words read correctly within 45 s is recorded. Psychometric properties are good, with most alternate forms and test–retest reliability coefficients at or above .90 in this age range (Torgesen et al., 2001). The standard score was utilized. The Texas Middle School Fluency Assessment (University of Houston, 2008) has subtests of Passage Fluency (PF) and Word Lists (WL), and was developed for the parent project. PF are 1-min probes of narrative and expository text. Each student read a set of five randomly selected probes across a spaced range of difficulty of approximately 500 Lexile points. These measures show good criterion validity ($r = .50$) with TAKS performance. The score used for the present work was the average number of linearly-equated words correctly read per minute across the five passages. WL are also 1-min probes, but of individual words. Each student read a set of three randomly selected probes within conditions across difficulty levels, with difficulty determined by word length and frequency parameters. WLs were constructed both from the PF texts, as well as from standard tables (Zeno, Ivens, Millard, & Duvvuri, 1995). Conditions included WL constructed from the PF that the student read, WL constructed from the PF that other students read, or WL constructed from standard tables. Criterion validity with TAKS is good ($r = .38$).

Comprehension—Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2002) Reading Comprehension subtest. The GRADE is a diagnostic reading test specifically for students in Pre-K-high school. It is a group-based, norm-referenced untimed test. For Reading Comprehension, participants read one or more paragraphs and answer multiple choice questions. The questions are designed to require metacognitive strategies, including questioning, predicting, clarifying, and summarizing. In a related subsample of Grade 6 students, internal consistency was found to be .82 (Vaughn, Cirino, et al. 2010a). The dependent measure analyzed was a prorated Comprehension composite standard score based on this measure alone (where the composite is typically based on both the Reading Comprehension and Sentence Comprehension subtests of the GRADE). The Passage Comprehension subtest of the WJ-III (see above) assesses reading comprehension at the sentence level using a cloze procedure; participants read the sentence or short passage and fills in missing words based on the overall context. The standard score was utilized. The Comprehension portion of the Texas Middle School Fluency Assessment (see above) was also utilized. This measure is related to the PF subtest, although the Comprehension portion required students to read the entire passage. Comprehension is assessed by asking a series of implicit and explicit multiple choice questions following reading of each passage. For purposes of the present study, the measure utilized was the proportion of questions correctly answered summed across the five passages read by the student.

Combined measures—A pre-publication version of the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010), then known as the Test of Sentence Reading Efficiency (TOSRE) is a 3-min assessment of reading fluency and comprehension. Students read short sentences and answer True or False. The raw score is the number of correct minus incorrect sentences. Criterion related validity with TAKS is good ($r = .56$). AIMSweb Reading Maze (Shinn & Shinn, 2002) is a 3-min, group-administered curriculum based assessment of fluency and comprehension. Students read a text passage; after the first sentence, in the place of seventh word thereafter students choose which of three words best fits the context of the story. The raw score is the number of targets correctly identified minus the number incorrectly answered within the time limit, which is convergent with raw scores computed in other manners (Pierce, McMaster, &

Deno, 2010). At each grade level, 15 different stories are available, with the particular story read randomly determined. The Maze subtest shows good reliability and validity characteristics. The Test of Silent Contextual Reading Fluency (TOSCRF; Hammill, Wiederholt, & Adam, 2006) assesses silent reading by having students read text graduating in complexity from simple sentences to passages increasing in length, grammar, and vocabulary. All words are capitalized with no space in between punctuation or consecutive words, and the task is to segment the passage appropriately. The TOSCRF shows good psychometric characteristics, and is appropriate for this age range. The standard score is the measure utilized.

Analyses

Prior to analyses, data was quality controlled in the field, after collection, and through standard verification features (basal and ceiling errors, out of range values). Univariate distributions were evaluated with both statistical and graphical techniques in the entire sample. They were also evaluated within struggling and typical reader subgroups because this distinction dichotomizes an underlying continuous distribution. Results were similar across groups. For several variables, there were some students who produced low values (−3SD), although the number of such students was always below 2 % (usually well below), but no distributions were severely nonnormal even with all values represented.

Primary analyses proceeded in two stages. The first stage involved separate (within group) and multi-group confirmatory factor analysis. The second stage involved comparing portions of struggling readers as identified by the state test with difficulty in decoding, fluency, and/ or comprehension as determined with nationally standardized measures. Each stage is elaborated.

Stage 1: MultiGroup confirmatory factor analysis (MGCFA)—MGCFA is a structural equation modeling technique that can be utilized to evaluate invariance of model parameters across groups. Comparing groups on composite means is possible utilizing standard techniques (e.g., ANOVA) and comparing groups on latent means is possible within the structural equation modeling framework (e.g., multiple-indicator multiple-cause [MIMIC] models). However, MCCFA has the advantage of being able to compare groups on additional measurement parameters as well, including latent factor loadings (configural or metric/weak invariance), intercepts/thresholds (strong invariance), and residual variances (strict measurement invariance) (Meredith, 1993), though differing terminology can be found in the literature.

Further, these graduated models can be tested in nested fashion at the global level (e.g., Chi-square comparisons), but individual parameters can be freed or fixed to more rigorously test the origin of the group differences. Our hypothesis expects that groups are invariant with regard to the way that observed indicators map onto their latent causes. By testing this invariance at each successive level described, we can evaluate the extent to which these latent factors have similar meaning in the two groups.

In evaluating the measurement invariance between the two groups we adopted a *modeling* rationale (as opposed to a *statistical* rationale based solely on χ^2 differences between nested models) for evaluating the mean and covariance structures which allowed us to evaluate the models using practical fit indices (e.g., comparative fit index or CFI, root mean square error of approximation or RMSEA; Little, 1997). We used this approach because of the relatively large number of constrained parameters in our models and the large sample size. The χ^2 statistic is sensitive to these factors. Consistent with the modeling rationale, we evaluated different levels of constrained models (i.e., more measurement invariance) for overall acceptable model fit (e.g., CFI > .95, RMSEA < .06, SRMR < .08, Hu & Bentler, 1999),

minimal differences between freer and more constrained models, uniform and unsystematic distribution of misfit indices for constrained parameters, and more meaning and parsimony in the constrained model than in the unconstrained model (Little, 1997). We also evaluated ΔCFI, which has been shown to be a robust indicator of measurement invariance (i.e., ΔCFI smaller than -0.01, where ΔCFI = CFI$_{constrained}$ $-$ CFI$_{unconstrained}$, indicates the more constrained model is acceptable; Cheung & Rensvold, 2002).

Stage 2: Criteria for external validity measures—A second purpose of this study was to determine whether students identified with reading comprehension difficulties on the state test would in fact exhibit difficulties not only on measures of comprehension, but also on measures of decoding and/or fluency. To investigate this hypothesis, we employed criteria external to TAKS for establishing difficulties in decoding, fluency, and reading comprehension. The decoding criteria was the 25th percentile (SS = 90) or below on the WJ-III Letter Word Identification subtest. Fluency criteria was the 25th percentile (SS = 90) or below on the TOWRE Sight Word Efficiency Subtest. Because we were particularly interested in reading comprehension, this criteria was evaluated in two ways, both representing the 25th percentile (SS = 90) or below, using the WJ-III Passage Comprehension subtest and/or the GRADE Reading Comprehension subtest. We also evaluated performance on the combined measures of comprehension and fluency, with the choice of measure determined from factor analytic results. We chose to use individual indicators given the use of these measures in practice; we did so with the expectation that they would load strongly on their respective factors, and because they are also nationally normed measures.

Results

Means and standard deviations, by reader group (struggling and typical) are provided in Table 2. As expected, typical readers outperformed struggling readers on these variables.

Hypothesis 1: Confirmatory factor analysis

The first hypothesis focused on the latent skills of decoding, fluency, and comprehension, with additional components assessing (a) the potential relevance of a separate construct involving measures that assess both fluency and comprehension, and (b) the structure of the latent constructs across groups.

First, we ran CFA models in each group separately. These models initially included nine indicator variables (two for decoding, four for fluency, three for comprehension); however, in all cases, the TOWRE PDE subtest seemed to be determined from both the decoding and the fluency factor; given its ambiguous position, and the fact that it was the only measure like this, it was deleted from further models. Table 3 displays fit statistics for the series of models with eight indicators of three latent domains, within each group. The first set of models in the table within each set progresses from least to most complex, but all involve the same participants and the same indicator variables and thus are nested. Model 1 within each group (Models 1a and 1b) are single factor models; model 2 (2a and 2b) separates comprehension measures from decoding and fluency factors; model 3 (3a and 3b) examines all three factors, and this model showed the best fit to these data, in both groups.

In the next set of analyses, the combined comprehension/fluency indicators are added. These measures are added first to the fluency factor (Models 4a and 4b), and then to the comprehension factor (Models 5a and 5b); these models are not nested, but other measures of fit clearly show preference for Models 5 over 4. However, when these indicators are added as a separate factor that combines fluency and comprehension (Models 6a and 6b), these are nested within Models 5 or 4, and are compared to Models 5 in Table 3. As shown,

Models 6 represented the best fit to the data. The fit of each of these Models 6 was improved with the addition of correlated errors (between WJ-3 Passage Comprehension/GRADE in the struggling group, and between TOSCRF/AIMSweb Maze and Word List Fluency/ Passage Fluency in the typical group), labeled Models 6a-modified and 6b-modified. Similar correlated error terms could also have been added to the aforementioned models, but the pattern of results was no different. The schematic (because correlated errors are not included) final model representing either group is displayed in Fig. 1.

For the multigroup analyses, the tested models progressed from least to most restrictive concerning parameters in the two groups, with results presented in Table 4. Model 1 is the least restrictive model specifying that the latent means of all four factors in both groups are zero, but all other parameters (factor loadings, intercepts—the means of the indicator variables, residual variances of the indicators) are free to vary. This model produced a good fit to the data (see Table 4), as expected, because this initial multigroup model is an additive one; the Chi-square contributions and degrees of freedom come from Models 6a and 6b in Table 3. The next step was to compare the above model to one in which the factor loadings were also set to be equivalent in the two groups. This model (Model 2 in Table 4) produced a relatively poor fit to the data. Thus, the hypothesis that the factor loadings for the two groups were equivalent was not supported; from a practical perspective what this means is that the indicators contribute to the latent factors differentially in the two groups. Therefore, a series of models were tested to determine if any of the factor loadings were invariant. Thus, Models 3-6 fixed only decoding, fluency, comprehension, and comprehension/fluency factor loadings, respectively. As shown in Table 4, Models 3 (decoding, fixed) and 6 (comprehension/fluency, fixed) showed fit comparable to the original model (Model 1), and Models 4 (fluency, fixed) and 5 (comprehension, fixed) showed poor fit relative to the original model. Therefore, Model 7 fixed the factor loadings of both the decoding and comprehension/fluency factors, which yielded an overall fit that was not different from Model 1. Thus, we concluded that these two factors across groups are invariant with regard to their factor loadings.

Models 8 and 9 then fixed intercepts and residual variances, respectively, for these two factors, to be equal across groups. Finally, Model 10 fixed both the intercepts and residual variances of these two factors across groups. Whenever intercepts were fixed to be the same across groups, their respective latent means were allowed to vary. These parameters were not evaluated with regard to the comprehension and fluency factors, since these are more restrictive models. Model fit is presented in Table 4.

Using the criteria described above, the model in which decoding and comprehension/fluency are both fixed in terms of factor loadings, intercepts, and residual variances is a better model than the free model (e.g., good overall fit, RMSEA = .06, SRMR = .06, ΔCFI = .001, ΔSRMR = .024). The end result of all of the comparisons was that while decoding and comprehension/fluency appear to be invariant in struggling versus typical readers, comprehension and fluency do not appear to be invariant. Practically, what these results imply is that these latter constructs are manifested differentially by the indicator variables studied here.

Model 10 in Table 4 represents the final multigroup model, with factor loadings, intercepts, and residuals all fixed across groups for decoding and comprehension/fluency factors, and all free across groups for comprehension and fluency factors (with one correlated residual variance in the struggling group, and two correlated residuals in the typical group). In this model, the standardized factor loading for the decoding factor was WJ-3 Letter Word Identification (.97 in both groups). For the fluency factor, loadings varied according to group—in struggling readers, all measures loaded .85 to .92, whereas in typical readers, the

strongest factor loading was for Word List Fluency (.98), and much lower for TOWRE SWE (.59) and Word List Fluency (.67). For the comprehension factor, WJ-3 Passage Comprehension had the strongest factor loading in both groups (.94 strugglers, .82 typical). For the Comprehension/Fluency factor, loadings were .80 for TOSRE, .73 for AIMSweb, and .64 for TOSCRF. Practically speaking, TOWRE SWE and Word List Fluency are not as reliable indicators of fluency for Typical readers as for Struggling readers whereas WJ-3 Passage Comprehension is a less reliable indicator of comprehension for Struggling than Typical readers (although GRADE and TCLD are comparable if less reliable indicators of comprehension than WJ-3 Passage Comprehension for both groups).

Inter-factor correlations are presented in Table 5. We tested the relations of the latent variables to one another across reader groups in a model comparison framework by constraining correlations among latent factors to be the same. Intercorrelations involving the comprehension/fluency factor, as well as the relation of decoding to fluency, could all be constrained to be equivalent across reader groups, $\chi^2 \Delta (df) = 10.748(4)$, ΔRMSEA = −.01, ΔCFI = .00, ΔBIC = −19.116 (model fit did not deteriorate). However, constraining the correlations of decoding with either fluency or comprehension did result in a substantially worse fitting model then the original, $\chi^2 \Delta (df) = 22.416(2)$, ΔRMSEA = +.02, ΔCFI = −.002, ΔBIC = +7.484; in the struggling reader group, the relation of decoding to fluency was stronger, and the relation of decoding to comprehension was weaker, than in typical readers. These comparative relations should be interpreted with caution given that the factors are composed in different manners between the two groups.

Hypothesis 2: Struggling reader subgroups

We denoted difficulties according to the normed indicators with the strongest factor loadings. The cut point was a standard score below the 25th percentile. The key variable for decoding was WJ-3 Letter Word Identification; for Fluency, TOWRE SWE; for Comprehension, WJ-3 Passage Comprehension (and GRADE); and for Comprehension/ Fluency, TOSRE. Given our focus on the comorbidity of different types of reading difficulties, and because the factors are constructed differentially in the two groups, we emphasized data for the struggling readers. This stage did not include 32 students (1.8 %) who were missing at least one of the key external measures, leaving 993 struggling readers. Among all struggling readers, 40 % exhibited difficulties in decoding, 39 % in fluency, 57 % (WJ-3) or 52 % (GRADE) in comprehension, and 67 % in comprehension/fluency. 18.4 % of students performed low on all five measures, 15 % on four measures, 15.1 % on three measures, 19.6 % on two measures, 17 % on one measure, and 14.8 % did not have any norm-referenced difficulties.

Table 6 presents results in the sample of struggling readers with at least one identified area of weakness (n = 846), without regard to overlap. Of these, 711 (84 %) had difficulties in comprehension, whereas slightly less than half had identified weakness in decoding or in fluency; 79 % had difficulties in comprehension/fluency. Few students had isolated difficulties; the largest subgroup of these students had only comprehension difficulties (12 % of struggling readers).

In terms of overlap, the most relevant combinations are likely those among the 711 students with difficulties in comprehension (that is, students below criterion on TAKS, who were also below criterion on either of two norm-referenced measures of reading comprehension). Here, 267 of these 711 (37.5 %) did *not* have decoding or fluency difficulties; the remaining 62.5 % were below the adopted threshold for this estimate. These numbers include the 102 with isolated comprehension difficulties; the other 165 are students with weaknesses on measures of comprehension as well as the measure of comprehension/fluency. It should also be noted that comprehension difficulties were indexed by either of two comprehension

measures, and that different subgroups of students were identified according to performance on the WJ-3 relative to the GRADE—382 students (38.5 %) performed low on both measures, 282 on neither (28.4 %), but 187 (18.8 %) were low on GRADE only, and 142 (14.3 %) on WJ-3 only.

While not a focus of the present study, the performance of students with data on these indicators who met criteria on TAKS was also examined. Of these 694 students, 109 (15.7 %) had comprehension but not decoding nor fluency difficulties, 47 (6.8 %) had comprehension with decoding or fluency difficulties, and 52 (7.5 %) had decoding and/or fluency but not comprehension difficulties. If the combined comprehension/fluency measure is included, then 63 (9.1 %) had comprehension without other difficulties, 93 (13.8 %) had comprehension with additional difficulties, and 91 (13.1 %) did not have comprehension difficulties, but did have other difficulties. Thus, difficulties occurred at a moderate rate (30 % to 35 %), but where they did, combinatory difficulties were frequent.

Discussion

The present study sought to address several gaps in the literature by evaluating the overlap among different measures of reading skill among middle school students within a latent variable context, comparing the component structure in struggling readers versus typical readers, and by evaluating the specificity versus overlap of difficulties among struggling readers. A four-factor model best characterized the external measures in this study, with latent factors of decoding, fluency, comprehension, and a fourth factor of combined comprehension and fluency representing timed measures with a comprehension component. We hypothesized that these factors would be invariant by reader groups, but found this to be true only for decoding and comprehension/fluency; in contrast, indicator loadings for fluency and comprehension were substantially different between the struggling reader and typical reader groups. Among struggling readers, students showed overlap in terms of the kinds of difficulties experienced; while comprehension difficulties were common, they overlapped considerably with decoding and/or fluency difficulties, which are powerful factors in determining the availability of information during reading.

Reading components

The present results are consistent with previous factor analytic studies that show decoding and comprehension to be separable reading components (Kendeou et al.,2009a, b; Nation & Snowling, 1997; Pazzaglia et al., 1993; Vellutino et al., 2007), and studies finding that subgroups of students exhibit varying combinations of difficulties across those components (Brasseur-Hock et al., 2011). Findings from this study extend previous studies in several ways, including the focus on middle school students, the comparison of struggling and typical readers with a large enough sample to adequately determine invariance, and the inclusion of measures that combine fluency and comprehension that have not been frequently used in previous studies despite their common use in schools.

Substantially worse fits resulted when models were used that forced measures to load with either fluency or comprehension relative to models that treated them as a separate construct. Latent correlations of this fluency/comprehension factor with each of the other components (including decoding) were high, and similar to one another (range .70 to .86; see Table 5), and could also be constrained across reader groups. These results suggest that such measures are not simply measures of "comprehension" nor of "fluency". Because all the measures of combined comprehension and fluency used the same response format (timed pencil and paper measures), it is unclear whether similar results would be found if different response formats were used (e.g., oral analogues). Performance on such oral analogues would more closely resemble reading fluency performance. However, if response format was so

influential, then it would be expected that the measures of fluency/comprehension involving silent reading would load with measures of comprehension. However, as Kim et al. (2011) have demonstrated, at least in younger students, oral and silent fluency (whose study included some measures combining fluency and comprehension) may have different roles for comprehension.

We hypothesized that findings would demonstrate invariance with regard to the factors examined here, but this was only partially supported. Where invariance was found (for decoding and comprehension/fluency), it was strict invariance (i.e., intercepts, loadings, and error variances were the same for both groups). In contrast, for the other factors (fluency and comprehension), we were unable to demonstrate even weak invariance (e.g., even factor loadings differed between groups). As such, findings from this study suggest that comprehension and fluency measures yield different findings in struggling versus typical readers. However, it was not simply the case that typical readers have more differentiated skills than struggling readers, as the four factor structure held for both reader groups. The correlations of Table 5 suggest that fluency is more related to decoding in struggling readers, whereas it is more related to comprehension in typical readers, though these correlations are difficult to compare, since the way these factors are indexed varies across groups.

Types of difficulties in struggling readers

The second hypothesis evaluated the extent to which comprehension difficulties occur in isolation. The selection and evaluation of these components followed from the results of the factor analyses described above. Among struggling readers, rates of specific difficulty were rather rare (1–12 %), but were highest for difficulties only in comprehension (Table 6). Even considering the whole sample of struggling readers, 68.2 % had difficulty in more than one domain. These high rates of overlap occurred despite the fact that comprehension was assessed in multiple ways (and the individual measures identify different students), and other areas were not assessed with multiple measures. More overlap would be expected if multiple indicators or fluency or decoding were utilized. Such results also highlight the limitations of selecting students based on observed measures, although there is a gap between how empirical studies can identify students versus how students are identified practically, in the field.

There are several potential explanations for the apparent discrepancy between the high degree of overlap seen in this study relative to figures found in national reports (Biancarosa & Snow, 2006). One could argue that the present results are sample specific and not generalizable, or that the selection measure for struggling readers utilized here (TAKS) was not sensitive enough to capture enough readers with specific comprehension difficulties. We view either of these possibilities as remote. Regarding generalizability, although the present study is not definitive, the numbers from the present study are consistent with several other studies (Catts et al., 2003; Leach et al., 2003; Shankweiler et al., 1999; Torppa et al., 2007) in showing that among struggling readers, isolated reading comprehension difficulties, while not rare, also do not comprise the bulk of struggling readers, who commonly have additional (and sometimes only) difficulties in decoding and/or fluency or other reading skills. Regarding TAKS, from a construct validity perspective, this measure related well to the other comprehension measures utilized in this study (e.g., in exploratory analyses, it loaded consistently with the other reading comprehension measures rather than with the other measures). As shown in Table 2, students who did not meet criteria on TAKS had performances on the nationally normed comprehension measures at the 30th (WJ-III) and 21st (GRADE) percentiles, suggesting correspondence of comprehension difficulties. Word-level (decoding and fluency) performances were also each at approximately the 30th percentile, and combined comprehension/fluency was at the 13th percentile, which adds to the evidence that students selected on the basis of reading comprehension also tend to have

more basic reading difficulties. A third possibility for the discrepancy is that not all reported figures are based directly on sample or population data; such figures may be susceptible to biases regarding the extent to which one believes that word level instruction (e.g., multi-syllable word work, morphology, phonics) is too "low level" and detracts from "higher level" instruction associated with comprehension (e.g., comprehension strategies).

Limitations

Some limitations to the current study should be noted. Additional approaches including person-centered techniques including cluster or latent class analyses could have been used to subgroup students, according to their performance. Such an approach has been used previously (e.g., Morris et al., 1998), and while its use here would likely provide useful information, the evaluation of overlap using individually normed measures is more in line with the aims of this particular study, which was to evaluate overlap according to pre-defined reading components. In addition, such individual measures are frequently used in research and practice in order to determine areas of difficulty. A more comprehensive approach would also have included supporting language variables (e.g., phonological awareness, rapid naming, vocabulary), though as noted, our current focus was on specific reading measures. It was, however, the case, that when a measure of listening comprehension was added in supplementary models, that overall model fits were similar, it did not obviate the relationship of the specific reading components to one another, and it was still the case that the reading components were separable at a latent level and showed varying degrees of invariance across reading levels. However, integrating reading-related language skills, along with demographic and instructional factors, into the present results would likely be beneficial.

Implications for practice

The majority of our sample of middle school struggling readers not only exhibited difficulty reading for understanding, but also faced difficulties in more basic word level reading skills. This is not to imply that reading intervention with older students should only focus on word decoding and fluency skills, but rather the majority of students will require interventions that address several components of reading. An exclusive focus on comprehension strategies may benefit the relatively small subgroup of students without difficulties in additional areas, though the present results do suggest that even if comprehension difficulties are identified, evaluation of additional components would clearly hold benefit in terms of developing the most effective multi-component approach. It may be that within a given setting, some types of students may be more or less common, and for schools and students with strong reading instruction backgrounds, students with specific comprehension difficulties may comprise the bulk of students who need additional assistance. However, in our large and diverse sample, the rate of overlap among different types of difficulties was strong, and these results are likely to generalize to other settings, particularly larger school districts.

Finding that multiple reading component processes are evident for adolescent struggling readers is not new although the present study does further our understanding of how such components might be related to one another, and their overlap. There have been several reviews and meta-analyses of interventions for adolescent struggling readers (Edmonds et al., 2009; Gersten, Fuchs, Williams, & Baker, 2001; Mastropieri, Scruggs, & Graetz, 2003; Swanson, 1999; Swanson and Hoskyn, 2001; Vaughn, Gersten, & Chard, 2000). Even year-long interventions are not always robustly effective for struggling readers in the older age group, particularly for reading comprehension (e.g., Corrin, Somers, Kemple, Nelson, & Sepanik, 2008; Denton, Wexler, Vaughn, & Bryan, 2008; Kemple et al., 2008; Vaughn, Denton, et al. 2010b; Vaughn, Wanzek, et al. 2010c). Because reading comprehension involves general language/vocabulary skills and background knowledge in addition to

decoding and fluency, the extent to which these factors are employed (or are successful) may differ in struggling and typical readers. Vaughn, Denton, et al. (2010b) and Vaughn, Wanzek, et al. (2010c) noted that the older readers in their study came from high-poverty backgrounds and exhibited significantly low levels of understanding of word meanings, background knowledge, concepts, and critical thinking. These types of findings highlight the significant challenges faced when seeking to substantially improve comprehension skills in older struggling readers, particularly when comprehension difficulties arise from a variety of sources. Effective routes to improving reading comprehension include targeting a variety of texts, utilizing cognitive strategies, particularly when strategy instruction is explicit and overt (Fagella-Luby & Deshler, 2008). The present study would also implicate the need to focus on more basic processes, as needed in struggling readers, even if identified as having comprehension difficulties. Further study can also help elucidate whether there is some "sufficient" criterion in either decoding or fluency that is needed in order to benefit from specific reading comprehension strategies.

Conclusion

These results show that the majority of middle school students with reading difficulties demonstrate reading problems that include word level reading, fluency, and comprehension. Sources of reading difficulties in middle school students are diverse, supporting the development of interventions that integrate instruction in accuracy, fluency, and comprehension. Such interventions also permit teachers to differentiate according to different student needs. Finally, the results suggest that simple screenings of accuracy and fluency, along with the broad based measures that are typically used at the state or district accountability level, may be essential for pinpointing the sources of reading difficulties and the nature and level of intensity of intervention needed.

Acknowledgments

References

Aaron PG, Joshi RM, Williams K. Not all reading disabilities are alike. Journal of Learning Disabilities. 1999; 32:120–137. [PubMed: 15499713]

Adlof SM, Catts HW, Lee J. Kindergarten predictors of second versus eighth grade reading comprehension impairments. Journal of Learning Disabilities. 2010; 43:332–345. [PubMed: 20463282]

Adlof SM, Catts HW, Little TD. Should the simple view of reading include a fluency component? Reading and Writing: An Interdisciplinary Journal. 2006; 19:933–958.

Biancarosa, G.; Snow, CE. Reading next—A vision for action and research in middle and high school literacy: A report from Carnegie Corporation of New York. 2nd ed. Alliance for Excellent Education; Washington, DC: 2006.

Brasseur-Hock IF, Hock MF, Kieffer MJ, Biancarosa G, Deshler DD. Adolescent struggling readers in urban schools: Results of a latent class analysis. Learning and Individual Differences. 2011; 21:438–452.

Buly MR, Valencia SW. Below the bar: Profiles of students who fail state reading tests. Educational Evaluation and Policy Analysis. 2002; 24:219–239.

Calhoon MB, Sandow A, Hunter CV. Reorganizing the instructional reading components: Could there be a better way to design remedial reading programs to maximize middle school students with

reading disabilities' response to treatment? Annals of Dyslexia. 2010; 60:57–85. [PubMed: 20087694]

Catts HW, Adlof SM, Weismer SE. Language deficits in poor comprehenders: A case for the simple view of reading. Journal of Speech, Language, and Hearing Research. 2006; 49:278–293.

Catts, HW.; Hogan, TP.; Adlof, SM. Developing changes in reading and reading disabilities. In: Catts, HW.; Kahmi, AG., editors. The connections between language and reading disabilities. Lawrence Erlbaum Associates; Mahwah, NJ: 2005. p. 50-71.

Catts HW, Hogan TP, Fey ME. Subgrouping poor readers on the basis of individual differences in reading-related abilities. Journal of Learning Disabilities. 2003; 36:151–164. [PubMed: 15493430]

Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal. 2002; 9:233–255.

Corrin, W.; Somers, M-A.; Kemple, J.; Nelson, E.; Sepanik, S. The Enhanced reading opportunities study: Findings from the second year of implementation (NCEE 2009–4036). National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; Washington, DC: 2008.

Cutting LE, Scarborough HS. Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. Scientific Studies of Reading. 2006; 10:277–299.

Denton CA, Wexler J, Vaughn S, Bryan D. Intervention provided to linguistically diverse middle school students with severe reading difficulties. Learning Disabilities Research & Practice. 2008; 23:79–89. [PubMed: 22736893]

Deshler DD, Schumaker JB, Alley GB, Warner MM, Clark FL. Learning disabilities in adolescent and young adult populations: Research implications. Focus on Exceptional Children. 1982; 15:1–12.

Edmonds M, Vaughn S, Wexler J, Reutebuch C, Cable A, Tackett K, et al. A synthesis of reading interventions and effects on reading outcomes for older struggling readers. Review of Educational Research. 2009; 79:262–300. [PubMed: 20072704]

Espin CA, Deno SL. Curriculum-based measures for secondary students: Utility and task specificity of text-based reading and vocabulary measures for predicting performance on content-area tasks. Diagnostique. 1995; 20:121–142.

Faggella-Luby M, Deshler D. Reading comprehension in adolescents with LD: What we know; What we need to learn. Learning Disabilities Research & Practice. 2008; 23:70–78.

Francis DJ, Snow CE, August D, Carlson CD, Miller J, Iglesias A. Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. Scientific Studies of Reading. 2005; 10:301–322.

Fuchs, D.; Fuchs, LS.; Mathes, PG.; Lipsey, MW. Reading differences between low-achieving students with and without learning disabilities: A meta-analysis. In: Gersten, RM.; Schiller, EP.; Vaughn, S., editors. Contemporary special education research: Syntheses of the knowledge base on critical instructional issues. Lawrence Erlbaum Associates Publishers; Mahwah, NJ, USA: 2000. p. 81-104.

Gersen R, Fuchs L, Williams J, Baker S. Teaching reading comprehension strategies to students with learning disabilities: A review of research. Review of Educational Research. 2001; 71:279–320.

Gough PB, Tunmer WE. Decoding, reading, and reading disability. Remedial and Special Education. 1986; 7:6–10.

Graney SB, Shinn MR. The effects of R-CBM teacher feedback in general educational classrooms. School Psychology Review. 2005; 34:184–201.

Grigg, WS.; Daane, MC.; Jin, Y.; Campbell, JR. The nation's report card: Reading 2002 (No. NCES 2003-521). U.S. Department of Education; Washington, DC: 2003.

Hammill, DD.; Wiederholt, JL.; Allen, EA. Test of silent contextual reading fluency (TOSCRF). Pro-Ed; Austin, TX: 2006.

Hasbrouck J, Ihnot C. Curriculum-based measurement: From skeptic to advocate. Perspectives on Language and Literacy. 2007; 33:34–42.

Hock MF, Brasseur IF, Deshler DD, Catts HW, Marquis JG, Mark CA, et al. What is the reading component skill profile of adolescent struggling readers in urban schools? Learning Disability Quarterly. 2009; 32:21–38.

Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6:1–55.

Joftus, S.; Maddox-Dolan, B. Left out and left behind: NCLB and the American high school. Alliance for Excellent Education; Apr. 2003

Kemple, J.; Corrin, W.; Nelson, E.; Salinger, T.; Herrmann, S.; Drummond, K. The enhanced reading opportunities study: Early impact and implementation findings (NCEE 2008–4025). National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; Washington, DC: 2008.

Kendeou P, Savage RS, Van den Broek P. Revisiting the simple view of reading. British Journal of Educational Psychology. 2009a; 79:353–370. [PubMed: 19091164]

Kendeou P, van de Broek P, White MJ, Lynch JS. Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. Journal of Educational Psychology. 2009b; 101:765–778.

Kim Y, Wagner RK, Foster E. Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. Scientific Studies of Reading. 2011; 15:338–362. [PubMed: 21747658]

Kirby JR, Savage RS. Can the simple view deal with the complexities of reading? Literacy. 2008; 42:75–82.

Leach JM, Scarborough HS, Rescorla L. Late-emerging reading disabilities. Journal of Educational Psychology. 2003; 95:211–224.

Little TD. Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. Multivariate Behavioral Research. 1997; 32:53–76.

MacGinitie, WH.; MacGinitie, RK.; Maria, K.; Dreyer, LG.; Hughes, KE. Gates-MacGinitie reading tests, fourth edition (GRMT-4). Riverside; Itasca, IL: 2000.

Mastropieri M, Scruggs T, Graetz J. Reading comprehension instruction for secondary students: Challenges for struggling students and teachers. Learning Disability Quarterly. 2003; 26:103–116.

McGrew, KS.; Woodcock, RW. Woodcock-Johnson III technical manual. Riverside; Itasca, IL: 2001.

Meredith W. Measurement invariance, factor analysis and factorial invariance. Psychometrika. 1993; 58:525–543.

Morris RD, Stuebing KK, Fletcher JM, Shaywitz SE, Lyon GR, Shankweiler DP, et al. Subtypes of reading disability: Variability around a phonological core. Journal of Educational Psychology. 1998; 90:343–347.

Nation K. Reading skills in hyperlexia: A developmental perspective. Psychological Bulletin. 1999; 125:338–355. [PubMed: 10349355]

Nation, K. Children's reading comprehension difficulties. In: Snowling, MJ.; Hulme, C., editors. The science of reading: A handbook. Blackwell; Osford, UK: 2005. p. 248-266.

Nation K, Clarke P, Marshall CM, Durand M. Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? Journal of Speech, Language, and Hearing Research. 2004; 47:199–211.

Nation K, Snowling MJ. Assessing reading difficulties: The validity and utility of current measures of reading skill. British Journal of Educational Psychology. 1997; 67:359–370. [PubMed: 9376312]

National Assessment of Educational Progress. The nation's report card. National Center for Education Statistics; Washington, DC: 2009.

National Reading Panel. Report of the national reading panel: Teaching children to read—Reports of the subgroups. National Institute of Child Health and Human Development; Washington, DC: 2000.

Pazzaglia P, Cornoldi C, Tressoldi PE. Learning to read: Evidence on the distinction between decoding and comprehension skills. European Journal of Psychology of Education. 1993; 8:247–258.

Perfetti CA, Yang C, Schmalhofer F. Comprehension skill and word-to-text integration processes. Applied Cognitive Psychology. 2008; 22:303–318.

Pierce RL, McMaster KL, Deno SL. The effects of using different procedures to score maze measures. Learning Disabilities Research & Practice. 2010; 25:151–160.

Psychological Corporation. Wechsler individual achievement test. Author; San Antonio, TX: 1992.

Rimrodt, SL.; Lightman, A.; Roberts, L.; Denckla, MB.; Cutting, LE. Are all tests of reading comprehension the same?. Poster presented at the 33rd annual international neuropsychological society meeting; Feb. 2005

Sabatini JP, Sawaki Y, Shore JR, Scarborough HS. Relationships among reading skills of adults with low literacy. Journal of Learning Disabilities. 2010; 43:122–138. [PubMed: 20179307]

Shankweiler D, Lundquist E, Katz L, Stuebing KK, Fletcher JM, Brady S, et al. Comprehension and decoding: Patterns of association in children with reading difficulties. Scientific Studies of Reading. 1999; 3:69–94.

Share DL, Leikin M. Language impairment at school entry and later reading disability: Connections at lexical versus supralexical levels of reading. Scientific Studies of Reading. 2004; 8:87–110.

Shinn MR, Good RH, Knutson N, Tilly WD, Collins V. Curriculum-based reading fluency: A confirmatory analysis of its relation to reading. School Psychology Review. 1992; 21:458–478.

Shinn, MR.; Shinn, MM. AIMSweb training workbook: Administration and scoring of reading maze for use in general outcome measurement. Edformation; Eden Prairie, MN: 2002.

Snowling M, Hulme C. The nature and classification of reading disorders—A commentary on proposals for DSM-5. Journal of Child Psychology and Psychiatry. 2012; 53:593–607. [PubMed: 22141434]

Storch SA, Whitehurst GJ. Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. Developmental Psychology. 2002; 38:934–947. [PubMed: 12428705]

Swanson HL. Reading research for students with learning disabilities: A meta-analysis of intervention outcomes. Journal of Learning Disabilities. 1999; 32:504–532. [PubMed: 15510440]

Swanson HL, Hoskyn M. Instructing adolescents with learning disabilities: A component and composite analysis. Learning Disabilities Research & Practice. 2001; 16:109–119.

Texas Education Agency. [Retrieved May 15, 2012] Chapter 2. Texas assessment of knowledge of skills (TAKS). 2006a. from http://www.tea.state.tx.us/student.assessment/techdigest/yr0607/digest07-chap02[01].pdf

Texas Education Agency. [Retrieved May 15, 2012] Chapter 19. Performance Assessments used on TAKS/TAKS-I and SDAA-II. 2006b. from http://www.tea.state.tx.us/student.assessment/techdigest/yr0607/digest07-chap19[1].pdf

Texas Education Agency. [Retrieved May 15, 2012] Chapter 3. Assessments for students with disabilities. 2006c. from http://www.tea.state.tx.us/student.assessment/techdigest/yr0607/digest07-chap03[01].pdf

Torgesen, J.; Wagner, R.; Rashotte, C. Test of word reading efficiency. Pro-Ed; Austin, TX: 2001.

Torppa M, Tolvanen A, Poikkeus A, Eklund K, Lerkkanen MK, Leskinen E, et al. Reading development subtypes and their early characteristics. Annals of Dyslexia. 2007; 57:3–52. [PubMed: 17849214]

University of Houston. Texas middle school fluency assessment. Author; Houston, TX: 2008.

van den Broek P, Rapp DN, Kendeou P. Integrating memory-based and constructionist processes in accounts of reading comprehension. Discourse Processes. 2005; 39:299–316.

Vaughn SR, Cirino PT, Wanzek J, Wexler J, Fletcher JM, Denton CA, Barth AE, Romain M, Francis DJ. Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. School Psychology Review. 2010a; 39(1):3–21. [PubMed: 21479079]

Vaughn S, Denton CA, Fletcher JM. Why intensive interventions are necessary for students with severe reading difficulties. Psychology in the Schools. 2010b; 47:432–444. [PubMed: 21072127]

Vaughn S, Gersten R, Chard D. The underlying message in learning disabilities intervention research: Findings from research syntheses. Exceptional Children. 2000; 67:99–114.

Vaughn S, Wanzek J, Wexler J, Barth A, Cirino PT, Fletcher J, et al. The relative effects of group size on reading progress of older students with reading difficulties. Reading and Writing: An Interdisciplinary Journal. 2010c; 23:931–956.

Vellutino FR, Fletcher JM, Snowling MJ, Scanlon DM. Specific reading disability (dyslexia): What have we learned in the past four decades? Journal of Child Psychology and Psychiatry. 2004; 45:2–40. [PubMed: 14959801]

Vellutino FR, Tunmer WE, Jaccard JJ, Chen R. Components of reading ability: Multivariate evidence for a convergent skills model of reading development. Scientific Studies of Reading. 2007; 11:3–32.

Wagner, RK.; Torgesen, JK.; Rashotte, CA.; Pearson, NA. Test of silent reading efficiency and comprehension. Pro-Ed; Austin, TX: 2010.

Wiederholt, JL.; Bryant, BR. Gray oral reading test-3. Pro-Ed; Austin, TX: 1992.

Williams, KT. Teacher's scoring and interpretive manual. American Guidance Service, Inc.; Circle Pines, MN: 2002. The group reading assessment and diagnostic evaluation (GRADE).

Wolf M, Bowers PG. The double-deficit hypothesis for the developmental dyslexias. Journal of Educational Psychology. 1999; 91(3):415–438.

Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson tests of achievement. 3rd ed. Riverside Publishing; Itasca, IL: 2001.

Zeno, SM.; Ivens, SH.; Millard, RT.; Duvvuri, R. The educator's word frequency guide. Touchstone Applied Science Associates; Brewster, NY: 1995.

Ziegler JC, Goswami U. Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. Psychological Bulletin. 2005; 131:3–29. [PubMed: 15631549]
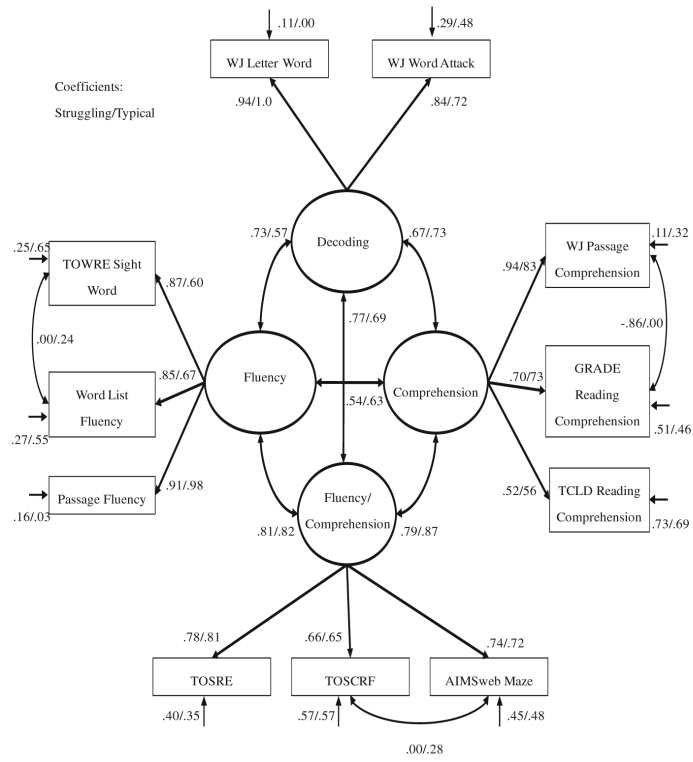
Fig. 1.
Schematic model for confirmatory factor analysis

Table 1

Demographic characteristics for the entire sample, and by reader group

Measure	Struggling	Typical	Total
Age (in years)	12.44 (1.02)	12.24 (0.96)	12.36 (1.00)
K-BIT 2 Verbal Knowledge	85.37 (12.8)	99.09 (13.1)	91.04 (14.57)
K-BIT 2 Matrices	94.02 (13.3)	104.34 (13.7)	98.55 (14.42)
K-BIT 2 Composite	88.20 (11.9)	102.60 (12.8)	94.52 (14.22)
Sex (female)	45.37	55.88	49.71
Limited English proficient	18.24	4.70	12.07
English as second language	14.54	1.52	9.15
Reduced/free lunch status	72.98	51.87	64.24
Special education	22.54	2.77	14.36
Retained	2.15	0.55	1.49
Ethnicity			
African American	40.68	38.73	39.87
Hispanic	42.34	27.94	36.38
Caucasian	14.44	28.49	20.25
Asian	2.34	4.70	3.32

K-BIT 2 = Kaufman Brief Intelligence Test-2; scores on K-BIT 2 are for descriptive purposes, expressed as standard scores. Verbal knowledge score is prorated based on vocabulary subtest (Riddles not given). Matrices score was administered at the end of the year, but is also age standardized. Age and K-BIT 2 display means (standard deviation). All other numbers are percentages within group; ethnicity percentages sum to one within rounding, within group. 58 students had scores of below 80 on both measures, 16 had scores below 75 (all struggling readers), and 7 had scores below 70 on both measures (all struggling readers). Age is in years, as of April 1, 2006 (though all standard scores derived from actual test data of evaluation). In terms of missing data, 19 students (1.09 %) missing limited English proficiency status, 43 students (2.46 %) missing English as second language status, 51 students (2.92 %) missing free lunch status, and 34 students (1.95 %) missing special education status. For continuous variables, reader groups differ, $p<.0001$, given the large sample sizes, though age differences are practically small. For categorical variables, reader groups differed on all variables, $p<.0001$

Table 2

Means and standard deviations on performance measures, by reader group

Measure	Struggling readers		Typical readers	
	N	*M*(SD)	*N*	*M*(SD)
TAKS	833	2027.87 (97.78)	723	2319.07 (138.95)
WJ-III Letter Word Identification	1,003	92.01 (12.3)	706	105.35 (12.0)
WJ-III Word Attack	1,002	95.50 (10.9)	706	103.69 (11.0)
TOWRE Sight Word Efficiency	1,001	92.38 (11.02)	706	101.98 (11.7)
TOWRE Phonemic Decoding Efficiency	1,001	94.53 (14.9)	706	105.36 (13.8)
Word List Fluency	1,021	72.99 (26.2)	719	93.23 (23.4)
Passage Fluency	1,020	112.06 (32.4)	719	145.59 (30.7)
WJ-III Passage Comprehension	1,003	86.19 (11.1)	706	98.98 (9.9)
GRADE Reading Comprehension	1,017	88.20 (9.9)	711	102.22 (12.2)
TCLD Reading Comprehension	1,017	73.65 (14.9)	718	86.66 (9.7)
TOSRE	1,025	83.34 (12.7)	720	99.12 (13.4)
AIMSweb Maze	1,019	13.17 (9.1)	718	22.92 (9.7)
TOSCRF	984	87.91 (10.9)	704	97.01 (10.6)

Total *N* is 1,748, including 1,025 struggling readers, and 723 typical readers. Missing data for any individual variable was less than 3 %, but all observations contributed data to confirmatory analyses. In the Struggling group, 192 students who received special education services did not take TAKS, but instead were assessed with the state determined alternative assessment (SDAA); these students were also considered struggling readers. TAKS is a standardized score, with a minimum pass score of 2,100. Measures of the WJ-III, TOWRE, GRADE, TOSRE, and TOSCRF are expressed in a traditional standard score metric (*M* = 100; SD = 15). Word List and Passage Fluency are number of words correctly read per minute; TCLD Reading Comprehension is the percent of items correctly answered across all five passages read by the student; AIMSweb Maze is the total raw score (correct-incorrect) within the 3 min limit

Table 3

Model fit (separate groups)

Model	χ^2 (*df*)	CFI	BIC	RMSEA (90 % CI)	SRMR	$\chi^2 \Delta$ (*df*)	*p*
Struggling readers							
Set I 1a	995.815 (20)	.795	53338.943	.218 (.207 to .230)	.090	–	–
Set I 2a	736.635 (19)	.849	53086.696	.192 (.180 to .204)	.072	259 (1)	.0001
Set I 3a	132.967 (17)	.976	52496.893	.082 (.069 to .095)	.042	604 (2)	.0001
Set II 4a	633.901 (41)	.910	74195.825	.119 (.111 to .127)	.071	–	–
Set II 5a	331.425 (41)	.956	73893.349	.083 (.075 to .092)	.047	–	–
Set II 6a	207.812 (38)	.974	73790.534	.066 (.057 to .075)	.038	113 (3)	.0001
Set II 6a-modified	166.316 (37)	.980	73755.970	.058 (.050 to .068)	.030	41 (1)	.0001
Typical readers							
Set I 1b	640.142 (20)	.753	37682.791	.207 (.193 to .221)	.086	–	–
Set I 2b	500.898 (19)	.808	37550.130	.187 (.173 to .202)	.073	139 (1)	.0001
Set I 3b	113.562 (17)	.962	37175.961	.089 (.074 to .104)	.040	387 (2)	.0001
Set II 4b	410.309 (41)	.905	52542.932	.112 (.102 to .122)	.056	–	–
Set II 5b	332.823 (41)	.925	52465.446	.099 (.089 to .109)	.050	–	–
Set II 6b	217.839 (38)	.954	52370.231	.081 (.071 to .092)	.040	114 (3)	.0001
Set II 6b-modified	149.150 (36)	.971	52314.690	.066 (.055 to .077)	.036	68 (2)	.0001

Models 1a and 1b—8 indicators of reading on 1 factor; models 2a and 2b—indicators of comprehension versus other measures; models 3a and 3b—separate decoding, fluency, and comprehension factors. All χ^2 values are significant, *p*<.0001. Within a set, models are compared to those preceding it

CFI comparative fit index, *BIC* Bayesian information criteria, *RMSEA* root mean square error of approximation, *SRMR* standardized root mean square residual

Table 4

Model fit (multigroup)

Model	χ^2 (df)	CFI	BIC	RMSEA (90 % CI)	SRMR	χ^2 Δ (df)	p	ΔCFI
1	315.466 (73)	0.977	126128.2	.062 (.055 to .069)	0.033	–	–	
	166.316S							
	149.150T							
2	458.865 (84)	0.964	126189.5	.071 (.065 to .078)	0.131	143 (11)	0.0001	–
	216.519S							0.013
	242.346T							
3	325.970 (75)	0.976	126123.8	.062 (.055 to .069)	0.048	10 (2)	0.04	–
	169.479S							0.001
	156.491T							
4	375.655 (76)	0.971	126166	.067 (.060 to .074)	0.095	60 (3)	0.0001	–
	181.051S							0.006
	194.604T							
5	375.703 (76)	0.971	126166	.067 (.060 to .074)	0.106	60 (3)	0.0001	–
	192.830S							0.006
	182.872T							
6	320.676 (76)	0.977	126111	.061 (.054 to .068)	0.044	5 (3)	ns	0.000
	168.493S							
	152.183T							
7	331.689 (78)	0.976	126107.1	.061 (.054 to .068)	0.054	16 (5)	ns	–
	171.801S							0.001
	159.897T							
8	355.477 (81)	0.974	126108.5	.062 (.056 to .069)	0.055	23 (3)	0.01	–
	192.932S							0.003
	162.545T							
9	363.673 (83)	0.973	126101.8	.062 (.056 to .069)	0.051	31 (5)	0.01	–
	187.293S							0.004
	176.380T							
10	386.363 (86)	0.971	126102	.063 (.057 to .070)	0.057	54 (8)	0.001	–
	210.075S							0.006
	176.288T							

Models 1 has no restrictions; Model 2 fixes all factor loadings to be equal across groups; Model 3 fixed only decoding factor loadings, Model 4 fixed only fluency factor loadings, Model 5 fixed only comprehension factor loadings, and Model 6 fixed only comprehension/fluency factor loadings; Model 7 fixes decoding and comprehension/fluency factor loadings, and freed comprehension and fluency factor loadings. Model 8 fixed the intercepts of the decoding and comprehension/fluency factors, Model 9 fixed the residual variances of the indicators of these factors, and Model 10 fixed both the intercepts and residual variances. All χ^2 values are significant, $p<.0001$. χ^2 Δ values in Models 2 through 7 are relative to Model 1; those in Models 8, 9, and 10 are relative to Model 7

CFI comparative fit index, *BIC* Bayesian information criteria, *RMSEA* root mean square error of approximation, *SRMR* standardized root mean square residual

Table 5

Factor intercorrelations among reader groups

Factor	Decoding	Fluency	Comprehension	Comprehension/fluency
Struggling readers				
Decoding	1.00			
Fluency	.703	1.00		
Comprehension	.666	.550	1.00	
Comprehension/fluency	.760	.812	.803	1.00
Typical readers				
Decoding	1.00			
Fluency	.590	1.00		
Comprehension	.741	.622	1.00	
Comprehension/fluency	.699	.807	.862	1.00

Indicators of decoding were WJ-3 Letter Word Identification and word attack; indicators of fluency are TOWRE SWE, WL fluency, and passage fluency; indicators of comprehension are WJ-3 passage comprehension, TCLD passage comprehension, and GRADE reading comprehension; indicators of comprehension/fluency were TOSRE, AIMSweb mazes, and TOSCRF

Table 6

Classification of reading difficulty

All struggling readers ($N = 993$)		
No difficulties	147	14.8 %
Some difficulties	846	85.2 %
Struggling readers with difficulty ($N = 846$)		
By area		
Decoding	399	47.16 %
Fluency	388	45.86 %
Comprehension/fluency	666	78.72 %
Comprehension	711	84.04 %
By specificity		
Decoding only	7	0.83 %
Fluency only	16	1.89 %
Comprehension/fluency only	68	8.04 %
Comprehension only	102	12.06 %

Proportions are given using WJ-III Passage Comprehension or GRADE Reading Comprehension