



# HHS Public Access

Author manuscript

*J Educ Psychol.* Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:

*J Educ Psychol.* 2015 August ; 107(3): 884–899. doi:10.1037/edu0000026.

## Examining General and Specific Factors in the Dimensionality of Oral Language and Reading in 4th–10th Grades

Barbara R. Foorman, Sharon Koon, Yaacov Petscher, Alison Mitchell, and Adrea Truckenmiller

Florida State University

### Abstract

The objective of this study was to explore dimensions of oral language and reading and their influence on reading comprehension in a relatively understudied population—adolescent readers in 4th through 10th grades. The current study employed latent variable modeling of decoding fluency, vocabulary, syntax, and reading comprehension so as to represent these constructs with minimal error and to examine whether residual variance unaccounted for by oral language can be captured by specific factors of syntax and vocabulary. A 1-, 3-, 4-, and bifactor model were tested with 1,792 students in 18 schools in 2 large urban districts in the Southeast. Students were individually administered measures of expressive and receptive vocabulary, syntax, and decoding fluency in mid-year. At the end of the year students took the state reading test as well as a group-administered, norm-referenced test of reading comprehension. The bifactor model fit the data best in all 7 grades and explained 72% to 99% of the variance in reading comprehension. The specific factors of syntax and vocabulary explained significant unique variance in reading comprehension in 1 grade each. The decoding fluency factor was significantly correlated with the reading comprehension and oral language factors in all grades, but, in the presence of the oral language factor, was not significantly associated with the reading comprehension factor. Results support a bifactor model of lexical knowledge rather than the 3-factor model of the Simple View of Reading, with the vast amount of variance in reading comprehension explained by a general oral language factor.

### Keywords

reading comprehension; oral language; decoding; adolescent literacy

---

The objective of the current study was to explore the dimensions of oral language and reading and their influence on reading comprehension in a relatively understudied population—adolescent readers in fourth through 10th grades. Reading in the narrow sense refers to conversion of written forms to oral language forms and, in a broader sense, it refers to obtaining meaning from print. Thus, learning to read implies understanding written language to approximately the same level as understanding oral language (e.g., Rayner,

---

Correspondence concerning this article should be addressed to Barbara R. Foorman, Florida Center for Reading Research, Florida State University, 2010 Levy Avenue, Ste. 100, Tallahassee, FL 32310. bfoorman@fcrr.org.

The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, the Educational Testing Service, or Florida State University.

Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). Oral language is a system through which spoken words convey knowledge, thought, and expression. Oral language is often described as consisting of phonology (the structure of sound), syntax (word order and grammatical rules), morphology (word structure), semantics (meanings of words and phrases), and pragmatics (the social rules of communication). In considering the dimensionality of oral language as it maps to written language comprehension in adolescents, we focus our attention on the interrelations among syntax, morphology, and semantics (specifically, vocabulary)—all aspects of lexical knowledge. We presume that phonological awareness skills (i.e., the awareness of sound structure) and social language skills (i.e., pragmatics) have been mastered by the vast majority of adolescents and, therefore, do not examine them directly in this study.

Most studies of adolescent reading examine predictors of reading comprehension using observed measures to highlight the importance of decoding (see García & Cain, 2014, for a meta-analysis), vocabulary (e.g., Clarke, Snowling, Truelove, & Hulme, 2010; Cromley & Azevedo, 2007; Eason, Sabatini, Goldberg, Bruce, & Cutting, 2013), inferencing (e.g., Tighe & Schatschneider, 2014), working memory (e.g., Cain, Oakhill, & Bryant, 2004), and, occasionally, syntax (e.g., Catts, Adlof, & Weismer, 2006). Further, with the exception of Cromley and Azevedo (2007) who used path analysis, studies used multiple regression (Cain et al., 2004; Clarke et al., 2010; Eason et al., 2013; Tighe & Schatschneider, 2014) or ANOVA (Catts et al., 2006) to address relations among the studied outcomes.

A noted limitation of using observed measures in the study of individual differences is that all assessments produce scores that are measured with error. Measurement error impacts the reliability of scores from the assessments, influences the ability to precisely estimate relations among variables, and leads to attenuation bias (Nunnally & Bernstein, 1994). Moreover, using multiple measures of a hypothetical construct, such as vocabulary (e.g., expressive and receptive), is commonplace in individual differences research. It is possible that a significant effect or relation may be observed for one measure and not another, leading the researcher to generate a conclusion about the measure's utility or implications for theory when the multiple measures may in fact be indicative of one broader construct. The confluence of measurement error in observed scores and in using multiple measures for a hypothetical construct may lead to an overgeneralization of findings. A solution to the problem of measurement error and multiple-measure analyses is to estimate latent variables from the manifest variables in a confirmatory factor analysis and then to hypothesize and compare models of different structural relations among the latent variables rather than at the level of the individual manifest variables.

There have been studies examining the structural relations among oral language and reading components in primary grade readers (e.g., Foorman, Koon, Petscher, Mitchell, & Truckenmiller, In Press; Kendeou, van den Broek, White, & Lynch, 2009; Lonigan, Burgess, & Anthony, 2000; Storch & Whitehurst, 2002). However, no studies have examined oral language components of vocabulary, syntax, and morphology from a latent variable modeling perspective with adolescents. Accordingly, to examine the dimensions of oral language and reading that predict reading comprehension for adolescents, we tested several models. First we tested the unidimensional model that assumes students' scores on

reading and oral language measures are one dimension (e.g., Mehta, Foorman, Branum-Martin, & Taylor, 2005). Second, we tested the assumption behind many clinical language assessments of four correlated latent variables—syntax, vocabulary, decoding, and reading comprehension (e.g., Tomblin & Zhang, 2006). Third, we tested the three-factor model of the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), in which reading comprehension is the product of decoding and oral language. We added syntax to the language component of the Simple View model because of the paucity of research on the contribution of syntax to reading comprehension in this age group. Finally, we examined a bifactor model that simultaneously estimated (a) the unique existence of specific language factors of syntax and vocabulary, as well as decoding and reading comprehension, and (b) a general factor of oral language comprised of the manifest syntax and vocabulary measures. Lack of unique factors of syntax and vocabulary would provide some empirical support for Perfetti's (1999) Reading Systems Framework that includes syntax as part of the lexicon.

## READING AND ORAL LANGUAGE COMPONENTS AS UNIDIMENSIONAL

Are reading and oral language constructs so highly related that they should be considered one dimension, a dimension that might be called literacy? Or, are the constructs independent but correlated with each other? Instructional battles have been waged for decades over integrated, comprehensive, holistic approaches to teaching reading versus componential, skills-based approaches (Foorman, 1995; Rayner et al., 2001).

There is one study that addressed the question of the dimensionality of literacy from a multivariate, multilevel perspective—a study by Mehta et al. (2005). This was a longitudinal study of 1,342 students in Grades 1–4 in 127 classrooms. Confirmatory factor models were fit with single- and two-factor structures at the student and at the classroom level. Results supported a unidimensional structure for reading and spelling, with the contribution of phonological awareness diminishing over the grades. Writing was the least related to the literacy dimension and the most influenced by teacher effects. Oral language ability—consisting of receptive vocabulary and similarities and information subtests from an intelligence test—was separable from the literacy dimension at the student level but not at the classroom level. Because the current study is at the student level, we might expect to see reading and oral language ability as correlated but separable. Mehta et al. (2005) did not study the dimensionality of their oral language measures to see if second-order factors existed.

Thus, in the unidimensional model in this study, we examined structural relations among 11 observed measures of vocabulary, syntax, morphology, decoding, and reading comprehension to test the possibility that they all belong to the same latent construct.

## READING AND ORAL LANGUAGE COMPONENTS AS FOUR CORRELATED FACTORS

Most language and reading assessments include separate tests of vocabulary, syntax, decoding, and reading comprehension, although many linguists regard syntax as part of the lexicon (e.g., Bates & Goodman, 2001), whereas others (e.g., Pinker, 1998; van der Lely,

Rosen, & McClelland, 1998) regard the lexicon and grammatical rules as belonging to separate systems. To better understand the separability of semantics and syntax, Tomblin and Zhang (2006) tested the dimensionality of language measures with children in kindergarten, second, fourth, and eighth grades. They tested confirmatory factor analyses of language tests grouped according to modality (receptive vs. expressive) and domain (vocabulary vs. syntax, which they call sentence use). They found no support for the receptive-expressive dimension, but they found that a two-dimensional model of vocabulary and sentence use fit the data better than the single-dimensional model, although the single-dimensional model also fit well in the earlier grades. The factor correlations for the tests administered across all three grade levels showed a decreasing trend, from .93 in second grade, to .86 in fourth grade, to .78 in eighth grade. Tomblin and Zhang speculated that the increasing differentiation of the lexical and grammatical systems could be due to: a) the nativist perspective that syntactic growth will reach an asymptote by adolescence whereas the lexicon will continue to develop (Chomsky, 1965); b) a developmental perspective that a sentence-level system such as syntax and morphology will diverge as the language user matures (Bates & Goodman, 2001); and c) a confounding of vocabulary with sentence use for younger students because of lack of familiarity with the words in the sentence task.

In sum, in accord with many clinical assessments, we tested a four-correlated-factors model that posited separate latent variables of syntax, vocabulary, decoding, and reading comprehension.

## **READING AND ORAL LANGUAGE COMPONENTS AS THREE CORRELATED FACTORS**

A model of three correlated factors is consistent with the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990). According to the Simple View of Reading, reading is the product of decoding and oral language. Tunmer and Chapman (2012) clarified that the Simple View was never meant as a complete theory of reading and that decoding and oral language were comprised of subcomponent processes that were directly or indirectly related to other factors. The link between decoding and reading comprehension in the Simple View has a strong research foundation. Many researchers report a strong relation between decoding and reading comprehension (e.g., Katzir et al., 2006, report  $R^2$  values around .90), whereas others report much weaker associations (e.g., Berninger, Abbott, Vermeulen, & Fulton, 2006, report  $R^2$  values ranging from .10 to .72 depending on the measure of reading comprehension). From the 110 studies examined in their meta-analysis of the decoding–reading comprehension relation, García and Cain (2014) found a strong corrected correlation of .74 between decoding and reading comprehension. The nature of the assessments moderated this correlation such that the relationship was stronger when decoding was measured with real words rather than nonwords, and with narrative text rather than expository text. Narrative text may reduce demands on prior knowledge because the text structures are more familiar (Best, Floyd, & McNamara, 2008).

García and Cain (2014) also found that age and listening comprehension were significant moderators of the decoding–reading comprehension relationship. Only 10 of the 28 studies that measured listening comprehension had sufficient information to include them in the

meta-analysis, but level of listening comprehension was a significant negative predictor ( $\beta = -.63$ ) and moderator of the decoding–reading comprehension relation, explaining a significant amount of variance in the effect sizes ( $R^2 = .40$ ). The negative relation is in accordance with the Simple View of Reading that higher listening comprehension is associated with lower decoding–reading comprehension relations. The significant effect of age is also in line with the Simple View’s prediction that the relation between decoding and reading comprehension will decrease with increasing age ( $\beta < .50$ ).

Recently Tunmer and Chapman (2012) tested 122 7-year-olds and claimed to confirm the hypothesis that one component of oral language—vocabulary—affects decoding, thereby challenging the assertion of the Simple View of Reading that decoding and oral language make independent contributions to reading comprehension. Wagner, Herrera, Spencer, and Quinn (2014) demonstrated that Tunmer and Chapman (2012) had in fact misspecified their structural equation model and left out a required covariance between the two latent variables of decoding and language comprehension. When the correct model was rerun, Wagner et al. showed that the Simple View of Reading had an excellent fit to the data. However, the correctly specified Simple View model turned out to be equivalent to models that replaced the covariance with a direct effect from oral language to decoding as well as a direct effect from decoding to oral language. They argue that the only way to differentiate these models is to conduct longitudinal studies and intervention studies.

Three longitudinal studies with primary-grade children relevant to the Simple View are those conducted by Protopapas, Sideridis, Mouzaki, and Simos (2007); Verhoeven and van Leeuwe (2008); and Quinn, Wagner, Petscher, and Lopez (2014). Protopapas et al. (2007) conducted their study in Crete where children learn to read the highly consistent Greek orthography. They found that decoding had a negligible effect on reading comprehension when vocabulary was taken into account. In Verhoeven and Van Leeuwe’s (2008) study with 2,143 children in the Netherlands learning to read the much less consistent Dutch orthography, reading comprehension in first grade was explained by decoding and listening comprehension. However, earlier vocabulary predicted later reading comprehension, whereas listening comprehension did not. Quinn et al. (2014) employed latent change score modeling with their sample of 316 first through fourth graders and found that vocabulary at one time point had an effect on the growth of reading comprehension during a subsequent time point. In contrast, reading comprehension at one time point did not have a significant effect on growth in vocabulary at a subsequent time point.

There are randomized intervention trials that support the independence of decoding (e.g., Bus & van IJzendoorn, 1999; Vadasy & Sanders, 2010; Wolff, 2011) and of oral language to reading comprehension. Direct effects of vocabulary instruction show moderate effects on norm-referenced vocabulary tests but small transfer effects to reading comprehension (Elleman, Lindo, Morphy, & Compton, 2009). Likewise, morphological awareness training shows limited transfer to reading comprehension (Bowers, Kirby, & Deacon, 2010).

However, there is one randomized trial of language comprehension that did demonstrate a significant effect on reading comprehension, and that is the study by Clarke et al. (2010) with 160 eight- and 9-year-old students with average decoding fluency and poor reading

comprehension. These students were randomly assigned to one of three types of individualized intervention: a) text comprehension training; b) oral language training; and c) a combination of text comprehension and oral language training. After 30 hours of training, students in all three groups had significantly higher reading comprehension than the untreated control group. The oral language group, however, had significantly higher gains from the end of the intervention to the follow-up assessment 11 months later. Additionally, the oral language and the combined training groups had higher expressive vocabulary than the untreated group and this was a mediator in their improved reading comprehension.

In sum, there is empirical support for the Simple View of Reading, but most of the evidence comes from studies of children learning to read, with oral language conceptualized as vocabulary and listening comprehension. A neglected area of oral language in studies of the Simple View of Reading is syntax. There is one study, however, that examined the dimensionality of oral language with 863 children in kindergarten, first, and second grades and found that vocabulary and syntax tended to share variance in comprising the oral language factor, along with listening comprehension (Foorman et al., In Press). Furthermore, the oral language factor predicted reading comprehension in a comparable manner to decoding fluency.

Studies of the Simple View of Reading that include syntax are also rare with older readers. Nation, Clarke, Marshall, and Durand (2004) found that in addition to vocabulary and semantic processing deficits, poor comprehenders also had difficulties with grammatical understanding of sentences. When Catts et al. (2006) examined the oral language performance of 490 eighth graders identified as poor comprehenders, they found them scoring on average at the 20th percentile in receptive vocabulary and at the 30th percentile in grammatical understanding. Although performance at the 30th percentile does not indicate a major deficit in syntax, it is sufficient to disrupt comprehension of written language. These poor comprehenders were part of an epidemiological study that started when the students first started school (Tomblin et al., 1997). Analysis of their performance in kindergarten, second, and fourth grades revealed that their language comprehension scores were significantly lower than typical readers and poor decoders even in these earlier grades.

Thus, the Simple View's claim that reading comprehension is the product of decoding and oral language has strong support; however, there is disagreement about what comprises the oral language component. In the current study we tested whether the oral language factor consisted of observed measures of syntax and vocabulary and then tested the relations among the three correlated factors of oral language, decoding, and reading comprehension.

## **READING AND ORAL LANGUAGE COMPONENTS IN A BIFACTOR MODEL**

A logical follow-up to the three-correlated factors model discussed above is to test a bifactor model of oral language. A bifactor model may be viewed as conceptually similar to a factor model with a second-order factor over multiple first-order factors. In addition to this conceptual similarity, they have been found to be algebraically equivalent under certain factor loading constraints (Rijmen, 2010). In general, as in this study, they are not equivalent

but nested models (Chen, West, & Sousa, 2006). While both models include individual factors as well as an additional modeled factor, there are a number of important differences. A measurement model with a second-order construct views a higher-order latent variable as causing the first-order factors and tests whether residual variance exists in first-order factors which may be best captured by the second-order factor (Reise, Morizot, & Hays, 2007). Conversely, the bifactor model seeks to capture variance among the observed measures through both specific factors of measured traits (e.g., syntax and vocabulary) as well as with a general factor (e.g., oral language). Chen et al. (2006) compared the use of bifactor and second-order factor models in representing general constructs consisting of multiple highly related domains and found that the bifactor model presents several advantages. One such advantage is the interpretability of the relationship between the specific factors (i.e., syntax and vocabulary) and an outcome variable (i.e., reading comprehension), after controlling for the general factor (i.e., oral language).

Perfetti's (1999) lexical knowledge perspective can potentially be tested in a bifactor model of relations among decoding, oral language, and reading comprehension. Perfetti's (1999) Reading Systems Framework includes knowledge sources (linguistic, orthographic, and general knowledge), basic language and cognitive processes, and the interactions among them. These interactions take place within a cognitive architecture with pathways between perceptual and memory systems and limited processing resources. A central focus of Perfetti's framework is the lexicon, which connects the word identification and comprehension systems. The former requires high-quality linguistic and orthographic information to ensure efficient word recognition. The latter inputs word recognition into building a mental model of the text, that is, in comprehending the text. Perfetti emphasizes that word-text integration processes are central to comprehension because they reoccur with each phrase (Perfetti, Landi, & Oakhill, 2005; Perfetti & Stafura, 2014).

Thus, a bifactor model could lend support to Perfetti's (1999) perspective that it is integrated syntactic and semantic knowledge (i.e., lexical knowledge) interacting with word identification that builds text understanding. To lend support, results of the bifactor model would need to show that the general oral language factor, rather than the specific syntax and vocabulary factors, predicts reading comprehension.

## THE CURRENT STUDY

The objective of the current study was to examine dimensions of oral language and reading that influence reading comprehension in a relatively understudied population—adolescent readers in fourth through 10th grades. Furthermore, the current study employed latent variable modeling of decoding fluency, vocabulary, syntax, and reading comprehension so as to represent these constructs with minimal error and to examine whether residual variance unaccounted for by oral language can be captured by specific factors of syntax and vocabulary. In the current study, we measured decoding fluency by giving students 45 sec to read a list of nonwords to tap phonological decoding and a list of sight words to tap orthographic knowledge that progress in difficulty from single to multiple syllables. We measured the semantic and syntactic aspects of word knowledge through a receptive vocabulary task and expressive tests of grammatical usage. Because of our emphasis on

lexical knowledge, we did not employ a separate measure of listening to passages. In fact, Adlof, Catts, and Little (2006) report a correlation in fourth grade between the measure of receptive vocabulary used here—the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007)—and the Listening to Paragraphs subtest of the Clinical Evaluation of Language Fundamentals (CELF-3; Semel, Wigg, & Secord, 1995) that is moderately strong (i.e., 0.67). Therefore, the PPVT-4 can serve as a proxy for listening comprehension in the current study and alleviates the concern about extreme memory and attentional demands required by listening to passages comparable to those administered for reading comprehension in fourth through 10th grades.

Although the data in the current study were not longitudinal, replication of the factor structure at each of the seven grades lends support to the examination of dimensions of oral language and reading that affect reading comprehension. Finally, in addition to testing one-, three-, and four-factor models (with the three-factor model being the Simple View of Reading model), we tested a bifactor model. The bifactor model allowed for testing (a) whether vocabulary and syntax items represent a general oral language factor as well as conceptually distinct, specific factors and (b) how the general and specific factors relate to reading comprehension. We asked the following research questions: Does a one-factor, three-factor, four-factor, or bifactor model of oral language fit better when examining the dimensionality of measures of syntax, vocabulary, and decoding fluency in fourth through 10th grades? What are the unique effects of the best fitting oral language and decoding fluency factors in predicting latent reading comprehension in fourth through 10th grades?

## METHOD

### Participants

The 1,792 participating students were from 18 schools in two large urban school districts in the Southeast and received positive parental consent to participate. The percentages of students participating in the Free and Reduced Priced Lunch program in the 18 schools ranged from 21.5% to 100%, with a median of 59%. Six of the 18 schools were Title I schools. The data were collected over a 3-year period. The first year of data, collected during the 2010–11 school year, served as a year to pilot assessment procedures. Students participating in the first year were identified as Cohort 1. Data collection in Years 2 and 3 occurred during the 2011–12 and 2012–13 school years, resulting in Cohort 2 and Cohort 3 samples. Across all grades students were predominantly female (53.9%). Demographic characteristics indicated that 31.9% of students were White, followed by 27.2% Hispanic, 20.3% Black, 4.6% Asian, 4.2% Multiracial, .4% Native American, and .5% Other. Grade-based sample and demographic data are reported in Table 1.

We examined the percentage of students at each grade who were served under the federal limited English proficiency (LEP) status, as determined by low scores on the Comprehensive English Language and Learning Assessment (CELLA; ETS, 2005). The numbers were less than 10% in Grades 4 (0.37%), 6 (0%), 8 (0.86%, with 13.79% missing), 9 (5.04%), and 10 (0.82%). The percentages were higher in Grade 5 (19.63%, with 0% missing) and Grade 7 (11.37%, with 12.04% missing). Because the LEP status provides access to federal funds for English-as-a-Second Language (ESOL) services, missing codes



for (LEP) most likely mean that the students are not served. However, this is impossible to verify in the available state database.

## Measures

**Syntax**—The Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals-4 (CELF-4; Semel, Wiig, & Secord, 2003) was used to measure the ability to recall and produce sentence structures of varying length and syntactic complexity. Semel et al. (2003, p. 186) provide information on construct validity by referencing Slobin and Welsh's (1973) work that indicates that children translate sentences into their own language system and then repeat the sentences using their own language rules. These researchers pointed out that features of a child's language system can be obtained using imitation only if the stimulus sentence is long enough to tax the child's memory capacity. Semel et al. (2003, p. 245) acknowledge that working memory is inherent in many of the language subtests of the CELF-4 and include a table of intercorrelations of the language subtests with the CELF-4 Number Repetition-Backward subtest to show that the correlations are moderate across ages in their norming sample (i.e., 0.42 for Recalling Sentences and Repetition-Backward).

In the Recalling Sentences subtest, the examiner reads a sentence and the student must repeat that sentence verbatim. Sentences repeated with no errors receive three points, sentences with only one error receive two points, sentences with two or three repetition errors receive one point, and those with four or more errors receive a score of 0. The sentences gradually increase in terms of number of words and linguistic complexity. This subtest is discontinued after five consecutive scores of 0 and the total score is established by totaling all points prior to discontinuation. This subtest is designed for use with students ages 5–21. Cronbach's alpha reliability coefficients for the Recalling Sentences subtest range between .86 and .93.

The ability to recognize and correct syntactical errors in sentences was measured by the Grammaticality Judgment subtest (GJT) of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 2008). Individuals are orally presented sentences by an examiner with and without grammatical errors and the individual must indicate whether or not there are errors. The examinee is asked to fix any perceived errors in the sentence without changing its meaning. The Grammaticality Judgment subtest is designed for students ages 7–21, with internal consistency reliability ranging from .85 to .94. The GJT reports criterion-related validity through the relationship of its subtests to other oral language assessments. After correcting for variability between norm groups, the GJT correlates to the Listening Comprehension and Oral Expression Scales (Carrow-Woolfolk, 1995) Oral Composite score at .75.

**Vocabulary**—Receptive vocabulary was measured by the Peabody Picture Vocabulary Test-4 (PPVT-4; Dunn & Dunn, 2007) and two subtests from the Study Aid and Reading Assistant (SARA; Educational Testing Service, 2009). The PPVT-4 is designed for administration as early as 2½ years, up to age 90 and over. In the PPVT-4, students point to the picture, from a group of four pictures, that best represents a word spoken by the

examiner. Reported reliability of the PPVT-4 is very high, with internal consistency reliability ranging from .92 to .98. The PPVT-4 has correlations ranging from .80 to .83 with the Expressive Vocabulary Test (Williams, 2007) for individuals 7- to 24-years-old. Correlations with the CELF-4 range from .67 to .79.

The Vocabulary subtest of the SARA measures the ability to understand the meaning of individual words and their relationships to topical knowledge. For each given target word, the student must select one of three responses that provides a synonym (e.g., cold: frigid) or meaning associate (e.g., hot: burn) for the target word. There are 38 total possible items for this subtest and students are given up to 6 minutes to complete all items. If students are not finished with all items at 6 minutes, the subtest ends and students move on to the next subtest. Cronbach's alpha reliability coefficients for the Vocabulary subtest that have been established in pilot testing with a sixth through eighth grade sample have ranged between .86–.88 (Sabatini, Bruce, & Steinberg, 2013).

The SARA Morphological Awareness subtest focuses on derivational morphology by employing a target consisting of a prefix or suffix attached to a root word as in *pre* [prefix] + *view* [root] (Educational Testing Service, 2009). This focus was based on the recognition that the academic vocabulary of middle school text contains many derived words. In this measure, students are presented with a cloze item with three choices, all of which are derived forms of the same root word, and must select the one word that makes sense in the sentence. For example: she is good at many sports, but her \_\_\_\_\_ is basketball [*specialty*, *specialize*, *specialist*]. There are 32 total possible items for this subtest and students are given up to 7 minutes to complete all items. If students are not finished with all items at 7 minutes, the subtest ends. Cronbach's alpha reliability coefficients for the Morphological Awareness subtest that have been established in pilot testing with a sixth through eighth grade sample have ranged between .90–.91 (Sabatini et al., 2013).

Validity evidence for the SARA Vocabulary and Morphological Awareness measures is provided by evidence that these measures, although correlated ( $r = .74$ ), added unique variance in predicting 3,372 middle school students' performance on their state reading test (O'Reilly, Sabatini, Bruce, Pillarisetti, & McCormick, 2012).

**Decoding fluency**—Decoding fluency was measured by two forms (Forms A and B) of the Phonetic Decoding Efficiency (PDE) and Sight Word Efficiency (SWE) subtests of the Test of Word Reading Efficiency-2 (TOWRE-2; Wagner, Torgesen, & Rashotte, 2012). TOWRE-2 was designed to be administered to students ages 6–24. In this test, the examiner asks students to read nonwords and sight words aloud as quickly as possible within 45 seconds. The alternate-forms reliability coefficient ranges from .82-.94 for the SWE subtest and from .86–.95 for the PDE subtest. The average test–retest coefficients for different forms of the subtests exceeds .90. The TOWRE is highly related to other reading measures (Wagner et al., 2012): a) decoding, such as the Letter-Word Identification subtest of the Woodcock-Johnson III ( $r = .76$ ; Woodcock, McGrew, & Mather, 2001), reading fluency ( $r = .91$ ) on the Gray Oral Reading Test-4th ed. (GORT-4; Wiederholt & Bryant, 2001), Test of Silent Contextual Reading Fluency (TOSWRF; Hammill & Wiederholt, 2006;  $r = .75$ ),

and the Woodcock Reading Mastery Test–Revised (WRMT-R; Woodcock, 1987) Passage Comprehension ( $r = .88$ ).

**Reading comprehension**—Reading comprehension was measured by the Gates-MacGinitie Reading Test-4 (GMRT; MacGinitie, MacGinitie, Maria, & Dreyer, 2000) and the Florida Comprehensive Assessment Test 2.0 Reading (FCAT 2.0; Florida Department of Education, 2013). Levels 3 (Grade 3) through 10/12 (Grade 12) of the GMRT are designed to measure students' abilities to read and understand different types of prose. Multiple-choice questions require students to construct an understanding based on both literal understandings of the passages and on inferences drawn from the passages, in addition to determining the meaning of words. The K-R 20 reliability coefficient is .92 for the Comprehension subtest. The correlation between the GMRT Comprehension subtest and Iowa Test of Basic Skills (Hoover, Dunbar, & Frisbie, 2001) is .77 (Morsy, Kieffer, & Snow, 2010).

FCAT 2.0 Reading is designed to serve as a measure of reading comprehension in Grades 3–10. Multiple-choice questions are aligned to four reporting categories: vocabulary, reading application, literary analysis (fiction and nonfiction), and informational text/research process. Like the GMRT, questions require both literal and inferential understandings. Cronbach's alpha reliability coefficients for FCAT 2.0 Reading range between .89 and .93 for the spring administration (Florida Department of Education, 2013). Content and scoring validity evidence is provided by the publisher, including descriptions of the processes used to develop the assessments and results of dimensionality studies (Florida Department of Education, 2013). Second-order confirmatory factor analysis (CFA) results are provided as evidence that each grade-level assessment reflects a single underlying construct of reading achievement. Root mean square error of approximation (RMSEA) values ranged from .015 to .022, comparative fit index (CFI) values ranged from .98 to .99, and Tucker–Lewis index (TLI) values ranged from .98 to .99.

## Procedures

A planned missing data design was employed in Years 2 and 3 of the study in order to reduce the average amount of test time for study participants (Graham, Taylor, Olchowski, & Cumsille, 2006). Study measures were assigned to one of three forms, where Form C included all observed measures and Forms A and B were created so that each form included at least one of the multiple measures for each hypothesized latent construct. For example, our hypothesized construct of syntax was measured by the Grammatical Judgment task, which was included on Form A, and the Recalling Sentences task, which was included on Form B. Form C included both measures. The GMRT was administered to all students in Years 1 and 2 and 50% of students in Year 3. All students were administered the FCAT 2.0 Reading assessment as a part of statewide achievement testing. Students were randomly assigned to receive one of the three forms of measures. The advantage of using a planned missing data design is that it reduces the testing burden on students and produces data that are considered missing completely at random, in that missingness is completely unrelated to the study variables (Enders, 2010).

Students were individually administered the study measures in the fall (late September–November) each year and the outcome measures in late April or May. The CASL was not administered to students in Cohort 1. Administration time ranged between 32 and 65 minutes for Forms A and B for each grade, and between 79 and 120 minutes for students completing Form C. Due to the duration of testing, students were given short breaks between assessments and, when needed, students completed the form over two sessions.

Assessments within a form were administered to students in a variable order. Sample sizes by grade, cohort, and measure are provided in Table 2.

## Data Analysis

CFA was used to identify the measurement model that best explained the covariances among the observed measures. Three traditional factor models were specified including: a) a unidimensional model of oral language indicated by the 11 observed measures of syntax, vocabulary, decoding fluency, and reading comprehension (Figure 1a); b) a four-factor model with correlated factors of syntax, vocabulary, decoding fluency, and reading comprehension (Figure 1b); and c) a three-factor model with correlated factors of oral language, represented by syntax and vocabulary measures, decoding fluency, and reading comprehension (Figure 1c). In addition to these representations, we estimated a bifactor model which included individual factors of syntax, vocabulary, decoding, and reading comprehension, as well as a general oral language factor which was indicated by all of the syntax and vocabulary variables (Figure 1d).

Note that Figure 1d shows that both the Recalling Sentences (i.e., CELF) and Grammatical Judgment (i.e., CASL) tasks are caused by the latent syntax factor, as well as the oral language factor. Similarly, the PPVT and the SARA Vocabulary (V) and Morphology (M) tasks are caused by the latent vocabulary factor and the oral language factor. This bifactor model specification tests the extent to which the variance across the indicators of syntax and vocabulary skills measure a global construct of oral language, and whether residual variance unaccounted for by oral language can be captured by specific factors of syntax and vocabulary.

As part of the hypothesized model, note that the syntax, vocabulary, and oral language factors in Figure 1d do not correlate with each other, but do correlate with both decoding fluency and reading comprehension. This specification results in a set of orthogonal first-order latent variables with different degrees of generality (Gustafsson & Balke, 1993), and allows one to test the extent to which the residual factors of syntax and vocabulary may uniquely relate to comprehension after controlling for the general factor of oral language (Reise et al., 2007).

Model fit was evaluated using multiple indices, including chi-square ( $\chi^2$ ), CFI, TLI, RMSEA, and standardized root-mean-square residual (SRMR). RMSEA values below .08, CFI and TLI values equal to or greater than .95, and SRMR values equal to or less than .05 are preferred for an excellent model fit (Hu & Bentler, 1999). For nested models, the  $\chi^2$  difference test also was used to evaluate model fit. A significant  $\chi^2$  difference test indicates that the more constrained model (i.e., the model with more degrees of freedom) provides

significantly worse fit to the data than the less constrained model (i.e., the model with fewer degrees of freedom).

Following the measurement model specification, structural equation modeling (SEM) was used to test the extent to which the latent oral language and decoding fluency factors identified from the CFA predicted the latent representation of reading comprehension. Analyses were conducted separately by grade due to differences in the number of cohorts. Developmental scale scores were used for the GMRT and the FCAT. In the absence of developmental scale scores, standard scores were used for the CASL and PPVT-4. Total correct scores were used for the remaining CELF-4, TOWRE-2, and SARA measures due to the unavailability of other score types.

## RESULTS

### Preliminary Analyses

Sample statistics for each measure are presented in Tables 3 and 4, including correlations, means, standard deviations, minimums and maximums. All of the syntax, vocabulary, and decoding measures were moderately to highly correlated with the reading comprehension measures. Correlations within constructs (i.e., syntax, vocabulary, and decoding) were generally higher than across constructs (see Table 3).

Missing data were considered by cohort and then grade, due to differences in the mechanism generating the missing data. Students participating in Cohort 1 were administered all measures, except CASL. Approximately 0%–8.9% of data were missing across all measures in the Grade 6 Cohort 1, with the exception of the FCAT 2.0 measure, which had 39.6% missing data. While close to 40% of these students were missing FCAT 2.0 scores, only two students were missing both FCAT 2.0 and GMRT scores. Little's (1988) Test of Missing Completely at Random (MCAR) demonstrated that all missing data were considered MCAR [ $\chi^2(86) = 94.23, p = .26$ ]. Similarly, up to 10.9% of data were missing in the Grade 9 Cohort 1, with the exception of the FCAT 2.0 measure, which had 33.7% missing data. In this group, all missing data were found to be MCAR with  $p = .07$ . As described previously, the data collection design for Cohorts 2 and 3 was based on a planned missing data design using three forms. Consistent with the design, missingness ranged from 31.7%–43.1% across all grades for measures of syntax, vocabulary, and decoding. Between 8.8%–15.9% of Cohort 2 students and between 52.8%–58.2% of Cohort 3 students had missing GMRT outcome data, which is aligned to the random sampling method used to assign students to outcome testing on this measure. While FCAT 2.0 missing data ranged from 4.8% to 20.6% across all grades, only 2.6%–5.3% of each grade-level sample was found to have no outcome measure (i.e., no FCAT 2.0 or GMRT measure).

Data from the planned missing design were leveraged by using full information maximum likelihood (FIML) in the factor analytic and structural equation models. FIML is advantageous as it requires an assumption of data missing completely at random or missing at random, which holds in the present design. Further, FIML has been shown to produce parameter estimates and associated standard errors which are unbiased, even with large amounts of missing data (Enders, 2006).

## Measurement Model

Results from the measurement model analyses by grade level are provided in Table 5. In fourth through 10th grades, model results followed the same pattern in that the unidimensional model consistently provided the worst relative fit to the data, and the unrestricted bifactor model provided the best fit to the data. While the criterion-based fit indices (i.e., CFI, TLI, RMSEA, and SRMR) for the three-factor, four-factor, and bifactor models were all within the acceptable range for model fit, relative fit indices (i.e., AIC, Adjusted BIC) across the models and grades favored the bifactor model. In fourth grade, the AIC and BIC differences between the next best fitting model (i.e., the three-factor model) and the bifactor model was  $AIC = 47$  and  $BIC = 45$ . To contextualize this difference, Raftery (1995) reported that information criteria differences of 3–9 constitute a positive difference, 10–100 is a strong difference, and 100 is considered decisive. As is apparent in Table 5, the AIC and BIC difference values in fourth grade provide strong support for the bifactor model, which had strong criterion fit [ $\chi^2(30) = 51.07$ ,  $CFI = .99$ ,  $TLI = .97$ ,  $RMSEA = .05$  (90% CI = .03, .07)]. In a similar vein, the bifactor and next best fitting model in fifth grade (i.e., three-factor) differed by  $AIC = 45$  and  $BIC = 42$  in favor of the bifactor model [ $\chi^2(30) = 42.62$ ,  $CFI = .99$ ,  $TLI = .99$ ,  $RMSEA = .04$  (90% CI = .00, .06)]. This process was repeated for each grade, and for each comparison the AIC and BIC were consistently lower for the bifactor model (see Table 5). Strong differences in AIC and BIC values were found in all grades with the exception of 10th grade, where the difference was positive but not strong.

## Structural Analysis

The emergence of the bifactor model for characterizing oral language and correlations with decoding and reading comprehension served as the basis for constructing the structural equation model. The bifactor model (represented in Figure 1d without grade level) generally tested the direct effects of the specific oral language factors of syntax and vocabulary, as well as decoding and the general factor of oral language in predicting reading comprehension.

The general oral language factor, measured both by syntax and vocabulary indicators, significantly predicted reading comprehension in all grades, with standardized coefficient estimates ranging from 0.72 to 0.96 (see Table 6). Conversely, the specific vocabulary factor significantly predicted reading comprehension in only seventh grade. Similarly, the specific syntax factor significantly predicted reading comprehension in only one grade (i.e., fourth grade). While decoding fluency did not significantly predict reading comprehension in any grade, correlations between the decoding fluency factor and the reading comprehension factor ranged from .47 to .74. Across all grades, the amount of variance in reading comprehension that the latent factors of decoding fluency and oral language accounted for was between 72% and 99%.

By way of illustration, the bifactor model for eighth grade is depicted in Figure 2. The oral language factor, measured by latent measures of syntax and vocabulary, significantly predicted the latent variable of reading comprehension ( $\beta = .87$ ,  $p < .001$ ). However, the specific factors of syntax and vocabulary, and the decoding fluency factor did not

significantly predict the reading comprehension factor ( $\beta = .01$ ,  $\beta = .24$ , and  $\beta = .16$ ,  $p > .05$ , respectively). Together these latent variables accounted for 98% of the variance in reading comprehension. In addition, decoding fluency correlated positively and significantly with the oral language factor,  $r = .40$ ,  $p < .001$  and syntax,  $r = .89$ ,  $p < .05$ .

Finally, we explored replicating the bifactor model in fifth grade and seventh grade where 19.63% ( $n = 67$ ) and 11.37% ( $n = 34$ ) of the sample, respectively, were designated as Limited English Proficient (LEP). Because of the very small sample size, a test of measurement and structural invariance via a multiple group model was not warranted. Rather, we used a multiple indicators multiple causes (MIMIC) analysis, which is a special case of SEM that tests the impact of a covariate (i.e., LEP status) on the model. The latent factors of oral language, decoding fluency, and reading comprehension were each regressed on LEP status; however, the models in both grades failed to converge. Lack of model convergence could be due to the low sample size. Woods (2009) found that MIMIC models with fewer than 100 individuals could pose problems with the model. Though a small sample precluded more rigorous test of differences, we opted to conduct multiple  $t$  tests on the estimated factor scores to compare students with the LEP designation with those without the LEP designation. There were statistically significant effects after applying a linear step-up to correct for multiple hypothesis testing. Large effect sizes using Hedge's  $g$  were observed favoring the nonLEP students in both grades on all factors: a) for oral language in fifth grade,  $t(318) = 9.14$ ,  $p < .001$  ( $g = 1.28$ ), and in seventh grade,  $t(261) = 7.24$ ,  $p < .001$  ( $g = 1.33$ ); b) for reading comprehension in fifth grade,  $t(318) = 7.88$ ,  $p < .001$  ( $g = 1.10$ ), and in seventh grade,  $t(261) = 7.76$ ,  $p < .001$  ( $g = 1.42$ ); and c) for decoding fluency in fifth grade,  $t(318) = 5.63$ ,  $p < .001$  ( $g = 0.79$ ) and in seventh grade,  $t(261) = 4.57$ ,  $p < .001$  ( $g = 0.84$ ).

## DISCUSSION

The objective of the current cross-sectional study was to examine the dimensions of oral language and decoding fluency and their influence on reading comprehension in a relatively understudied population—adolescent readers in fourth through 10th grades—and to do so by estimating latent variables via confirmatory factor analysis and by comparing their structural relations in one-, four-, three-, and bifactor models.

### Comparing Measurement Models and Structural Relations of Oral Language and Decoding Fluency and Their Influence on Reading Comprehension in 4th–10th Grades

Confirmatory factor analyses showed that across all seven grades the unidimensional model consistently provided the worst fit to the data and the bifactor model consistently provided the best fit. The poor fit of the unidimensional model seems counter to the findings of Mehta et al. (2005), but remember that their findings of unidimensionality in literacy measures were at the classroom level and that, at the student level, oral language and reading were highly correlated but separable.

The second model we tested had four correlated factors of syntax, vocabulary, decoding fluency, and reading comprehension and represented the model most consistent with clinical assessments of reading and language. Tomblin and Zhang (2006) tested the assumption of

clinical language assessments that syntax and vocabulary are separate factors and found that the dimensionality of latent factors for syntax and vocabulary went from a single dimension in kindergarten and second grade to two dimensions in fourth and eighth grades based on declining factor correlations. Data from the current study do not support increasing differentiation of the general oral language factor with grade beyond the unique associations of syntax and vocabulary with reading comprehension in one grade each. Findings from the current study suggest cautious interpretation of studies that present vocabulary (e.g., Clarke et al., 2010; Cromley & Azevedo, 2007; Eason et al., 2013; Ouellette & Beers, 2010) or syntax (e.g., Nation et al., 2004; Catts et al., 2006) as sole oral language predictors of reading comprehension in fourth through 10th grades.

These data are also consistent with the emergent factor structure of oral language and its relation to comprehension that we found in kindergarten, first, and second grades (Foorman et al., In Press). In first and second grades, the oral language and decoding fluency factors equally and strongly predicted reading comprehension outcomes. The structure of oral language differed slightly in the two grades. In first grade, the dimension of oral language was a second-order factor of listening comprehension, syntax, and vocabulary. In second grade, oral language was a second-order factor consisting of vocabulary and what was shared between listening comprehension and syntax.

The model we tested with three correlated factors of decoding, oral language, and reading comprehension is the Simple View of Reading model (Gough & Tunmer, 1986; Hoover & Gough, 1990). We added to the Simple View a relatively understudied construct—syntax. We tested a bifactor model in order to test whether residual variance existed in the first-order factor of oral language which might best be captured by the second-order factors of syntax and vocabulary. Perfetti's (1999) Reading Systems Framework would argue that the second-order factors of syntax and vocabulary not only share variance but do not exhibit unique effects on reading comprehension.

The bifactor model fit the data best in all seven grades and explained 72%–99% of variance in reading comprehension, with an average of 91%. This finding lends strong support to the Simple View of Reading from a latent-variable, bifactor perspective consisting of factors of reading comprehension and decoding and an oral language factor comprised of specific factors of syntax and vocabulary that may uniquely relate to the reading comprehension factor. In the current study, the specific factors uniquely predicted reading comprehension only in one grade each (syntax in fourth grade and vocabulary in seventh grade). Thus, it is the general oral language factor consisting of second-order factors of vocabulary and syntax that explained a large amount of the variability in reading comprehension (with correlations ranging from .72 to .96) and not the specific factors themselves.

These results lend support to the Reading Systems Framework's integration of syntax and semantics into lexical knowledge, which, in turn, plays a central role in linking word identification (i.e., decoding) with comprehension processes (Perfetti, 1999; Perfetti et al., 2005; Perfetti & Stafura, 2014). One could argue that our operationalization of syntax as sentence use and vocabulary as receptive vocabulary (PPVT-4), synonyms and meaning associates (SARA Vocabulary), and derivational morphology (SARA Morphological



Awareness) biases our measurement of oral language toward lexical knowledge rather than text-based knowledge and thus predisposes results that favor Perfetti's framework. However, as pointed out earlier, PPVT-4's moderately strong correlation with listening comprehension passages does give credence to it as a proxy for listening comprehension (Adlof et al., 2006). A more important caution to interpreting the current results as a complete test of Perfetti's (1999) Reading Systems Framework is that our measures of lexical knowledge do not include a productive orthographic measure such as spelling dictation task. The SARA Morphological Awareness task includes recognition of derived forms that involve shifts in phonology and orthography (e.g., *sign* to *signal* or *pity* to *piteous*), but a spelling dictation task could test the high quality lexical representations necessary to ensure efficient word recognition and construction of meaning units.

### Implications of a General Oral Language Factor

The current findings show that the vast proportion of variance in the reading comprehension factor in fourth–10th grades was accounted for by a general oral language factor. The decoding fluency factor did not make a unique contribution to reading comprehension in any grade. However, decoding fluency was significantly and moderately correlated with the general oral language factor and the reading comprehension factor in all grades, with factor correlations ranging from .40–.60 for the former and .47–.74 for the latter. The point is that in the presence of the oral language factor the decoding fluency factor was not significantly associated with the reading comprehension factor.

In their meta-analysis, García and Cain (2014) found a strong average, corrected correlation of .74 between decoding and reading comprehension, which declined with age. Sixty-one of the 110 studies in the meta-analysis included students over 10 years of age, comparable to the ages in the current study. Additionally, García and Cain (2014) found that listening comprehension was a significant moderator of the decoding-reading comprehension relation but that vocabulary was not. The current study's findings support and help explain the García and Cain (2014) results in that in the presence of the oral language factor, consisting of syntax and vocabulary, the decoding fluency factor was not significantly related to the reading comprehension factor and the specific vocabulary factor was not uniquely and significantly associated.

Thus, the García and Cain (2014) results and the current findings shed doubt on vocabulary's role as a proxy for the language comprehension component in the Simple View of Reading. Several studies in the primary grades found that print-related measures predicted reading outcomes in the primary grades but not oral language measures (Lonigan et al., 2000; Storch & Whitehurst, 2002; Schatschneider et al., 2004). However, other studies support oral language, including vocabulary, as predicting reading outcomes in the primary grades (e.g., Foorman et al., In Press; Clarke et al., 2010; Protopapas et al., 2007). In fact, as Wagner et al. (2014) point out, Tunmer and Chapman's (2012) attempt to study the direct and indirect effects of vocabulary actually supports the Simple View of Reading's claim that decoding and language comprehension both directly predict reading comprehension. Our current findings that the bifactor model best fit the data help clarify that vocabulary's primary role in predicting reading comprehension is as part of a general oral language factor

and not as a unique predictor that explains individual differences in reading comprehension. This important finding stands in contrast to those who suggest a major, unique role for vocabulary in predicting reading comprehension (Cromley & Azevedo, 2007; NCES, 2013; Ouellette & Beers, 2010; Quinn et al., 2014; Sabatini et al., 2013).

### Limitations

Caution in interpreting the results of our study is advised for several reasons. First, the sample sizes for each grade are relatively small for estimating such complex models. Second, missing codes on students' limited English proficient (LEP) status made it impossible to compare the bifactor model in non-LEP students versus LEP students, although the large effects using Hedges' *g* suggest pursuing measurement and structural invariance testing in a larger sample of English language learners. Third, results are based on cross-sectional rather than longitudinal samples. Fourth, the models are based on the measures used to inform constructs. Future studies might include a listening comprehension test in the oral language construct, a spelling dictation task as a written component of lexical knowledge, and a connected text reading fluency task to potentially expand the construct of decoding fluency (see Kim, Wagner, & Lopez, 2012). Fifth, most of our predictor variables were measured by conventional clinical tests administered individually to students, and the two measures of reading comprehension—both the Florida accountability test and the nationally normed test—consisted of expository and literary passages with multiple choice questions administered in a group format. Best et al. (2008) found that decoding explained less variance in reading comprehension in a multiple-choice task than in a free-recall task. Keenan and Betjemann (2006) caution that a large number of passage-dependent items may produce stronger relations between decoding and reading comprehension than a large number of passage-independent items. In sum, it is important to replicate the results of the current study in a large, longitudinal sample of native English speakers and English language learners with additional measures with varying formats representing different ways to conceptualize the constructs of oral language, decoding, and reading comprehension.

### CONCLUSION

One-, four-, three-, and bifactor models of oral language were compared in a sample of 1,792 adolescent readers in fourth through 10th grades. Results supported a bifactor model of oral language in all seven grades, with the vast amount of variance in reading comprehension explained by a general oral language factor. The one-factor model tested whether all 11 observed measures of oral language and reading belonged to the same latent variable. The four-factor model tested whether vocabulary, syntax, decoding fluency, and reading comprehension were separate latent variables as implied by many clinical assessments. The three-factor model was a latent variable version of the Simple View of Reading, according to which reading comprehension is the product of decoding fluency and language comprehension.

The current bifactor results challenge the Simple View of Reading's claim that decoding and language measures are unique predictors of reading comprehension. Vocabulary and syntax were not specific factors that, along with decoding fluency, predicted reading

comprehension. Instead, vocabulary and syntax were part of a general oral language factor which represented lexical knowledge (Perfetti, 1999; Perfetti & Stafura, 2014). The decoding fluency factor was correlated with reading comprehension and oral language factors in all grades, but, in the presence of the oral language factor, was not significantly associated with the reading comprehension factor.

## Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through a subaward to Florida State University from Grant R305F100005 to the Educational Testing Service as part of the Reading for Understanding Initiative.

## References

- Adlof S, Catts H, Little T. Should the simple view of reading include a fluency component? *Reading and Writing*. 2006; 19:933–958. <http://dx.doi.org/10.1007/s11145-006-9024-z>.
- Bates, E.; Goodman, J. On the inseparability of grammar and the lexicon: Evidence from acquisition. In: Tomasello, M.; Bates, E., editors. *Language development: The essential readings*. Oxford, UK: Blackwell; 2001. p. 134–162.
- Berninger VW, Abbott RD, Vermeulen K, Fulton CM. Paths to reading comprehension in at-risk second-grade readers. *Journal of Learning Disabilities*. 2006; 39:334–351. <http://dx.doi.org/10.1177/00222194060390040701>. [PubMed: 16895158]
- Best R, Floyd R, McNamara D. Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*. 2008; 29:137–164. <http://dx.doi.org/10.1080/02702710801963951>.
- Bowers P, Kirby J, Deacon S. The effects of morphological instruction on literacy skills. A systematic review of the literature. *Review of Educational Research*. 2010; 80:144–179. <http://dx.doi.org/10.3102/0034654309359353>.
- Bus A, van IJzendoorn M. Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*. 1999; 91:403–414. <http://dx.doi.org/10.1037/00220663.91.3.403>.
- Cain K, Oakhill J, Bryant P. Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*. 2004; 96:31–42. <http://dx.doi.org/10.1037/0022-0663.96.1.31>.
- Carrow-Woolfolk, E. *Oral and written language scales*. Circle Pines, MN: American Guidance Service; 1995.
- Carrow-Woolfolk, E. *Comprehensive assessment of spoken language*. Torrance, CA: Western Psychological Services; 2008.
- 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. [http://dx.doi.org/10.1044/1092-4388\(2006\)023](http://dx.doi.org/10.1044/1092-4388(2006)023).
- Chen FF, West SG, Sousa KH. A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*. 2006; 41:189–225. [http://dx.doi.org/10.1207/s15327906mbr4102\\_5](http://dx.doi.org/10.1207/s15327906mbr4102_5).
- Chomsky, N. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press; 1965.
- Clarke PJ, Snowling MJ, Truelove E, Hulme C. Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*. 2010; 21:1106–1116. <http://dx.doi.org/10.1177/0956797610375449>. [PubMed: 20585051]
- Cromley J, Azevedo R. Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*. 2007; 99:311–325. <http://dx.doi.org/10.1037/00220663.99.2.311>.
- Dunn, L.; Dunn, D. *Peabody Picture Vocabulary Test-4*. San Antonio, TX: Pearson; 2007.

- Eason SH, Sabatini J, Goldberg L, Bruce K, Cutting LE. Examining the relationship between word reading efficiency and oral reading rate in predicting comprehension among different types of readers. *Scientific Studies of Reading*. 2013; 17:199–223. <http://dx.doi.org/10.1080/10888438.2011.652722>. [PubMed: 23667307]
- Educational Testing Service. Comprehensive English language learning assessment CELLA): Technical summary report. Princeton, NJ: Author; 2005.
- Educational Testing Service. SARA: Study Aid and Reading Assistant. Princeton, NJ: Author; 2009.
- Elleman A, Lindo E, Morphy P, Compton D. The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*. 2009; 2:1–44. <http://dx.doi.org/10.1080/19345740802539200>.
- Enders, CK. Analyzing structural equation models with missing data. In: Hancock, GR.; Mueller, RO., editors. *Structural equation modeling: A second course*. Greenwich, CT: Information Age Publishing; 2006. p. 313-342.
- Enders, CK. *Applied missing data analysis*. New York, NY: Guilford Press; 2010.
- Florida Department of Education. Florida statewide assessments 2013 yearbook. Tallahassee, FL: Author; 2013.
- Foorman BR. Research on “The Great Debate”: Code-oriented versus whole-language approaches to reading instruction. *School Psychology Review*. 1995; 24:376–392.
- Foorman B, Herrera S, Petscher Y, Mitchell A, Truckenmiller A. The structure of oral language and reading and their relation to comprehension in kindergarten through grade 2. *Reading and Writing: An Interdisciplinary Journal*. (In Press).
- García JR, Cain K. Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*. 2014; 84:74–111. <http://dx.doi.org/10.3102/0034654313499616>.
- Gough P, Tunmer W. Decoding, reading, and reading disability. *Remedial and Special Education*. 1986; 7:6–10. <http://dx.doi.org/10.1177074193258600700104>.
- Graham JW, Taylor BJ, Olchowski AE, Cumsille PE. Planned missing data designs in psychological research. *Psychological Methods*. 2006; 11:323–343. <http://dx.doi.org/10.1037/1082-989X.11.4.323>. [PubMed: 17154750]
- Gustafsson JE, Balke G. General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*. 1993; 28:407–434. [http://dx.doi.org/10.1207/s15327906mbr2804\\_2](http://dx.doi.org/10.1207/s15327906mbr2804_2).
- Hammill, D.; Wiederholt, JL. *Test of Silent Contextual Reading Fluency*. Austin, TX: Pro-Ed; 2006.
- Hoover, HD.; Dunbar, SB.; Frisbie, DA. *Iowa Tests of Basic Skills (ITBS) Forms A, B, and C*. Rolling Meadows, IL: Riverside Publishing Company; 2001.
- Hoover W, Gough P. The simple view of reading. *Reading and Writing*. 1990; 2:127–160. <http://dx.doi.org/10.1007/BF00401799>.
- Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55. <http://dx.doi.org/10.1080/10705519909540118>.
- Katzir T, Kim Y, Wolf M, Kennedy B, Lovell M, Morris R. The relationship of spelling recognition, RAN, and phonological awareness to reading skills in older poor readers and younger reading-matched controls. *Reading and Writing*. 2006; 19:845–872. <http://dx.doi.org/10.1007/s11145-006-9013-2>.
- Keenan J, Betjemann R. Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading*. 2006; 10:363–380. [http://dx.doi.org/10.1207/s1532799xssr1004\\_2](http://dx.doi.org/10.1207/s1532799xssr1004_2).
- Kendeou P, van den 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*. 2009; 101:765–778. <http://dx.doi.org/10.1037/a0015956>.
- Kim YS, Wagner RK, Lopez D. Developmental relations between reading fluency and reading comprehension: A longitudinal study from grade 1 to grade 2. *Journal of Experimental Child Psychology*. 2012; 113:93–111. <http://dx.doi.org/10.1016/j.jecp.2012.03.002>. [PubMed: 22726256]

- Little RJA. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*. 1988; 83:1198–1202. <http://dx.doi.org/10.1080/01621459.1988.10478722>.
- Lonigan CJ, Burgess SR, Anthony JL. Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology*. 2000; 36:596–613. <http://dx.doi.org/10.1037/0012-1649.36.5.596>. [PubMed: 10976600]
- MacGinitie, W.; MacGinitie, R.; Maria, K.; Dreyer, L. *Gates-MacGinitie Reading Tests*. 4. Rolling Meadows, IL: Riverside Publishing; 2000.
- Mehta P, Foorman BR, Branum-Martin L, Taylor WP. Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1–4. *Scientific Studies of Reading*. 2005; 9:85–116. [http://dx.doi.org/10.1207/s1532799xssr0902\\_1](http://dx.doi.org/10.1207/s1532799xssr0902_1).
- Morsy, L.; Kieffer, M.; Snow, C. *Final report from Carnegie Corporation of New York’s Council on Advancing Adolescent Literacy*. New York, NY: Carnegie Corporation of New York; 2010. *Measure for measure: A critical consumers’ guide to reading comprehension assessments for adolescents*.
- 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. [http://dx.doi.org/10.1044/1092-4388\(2004\)017](http://dx.doi.org/10.1044/1092-4388(2004)017).
- National Center for Education Statistics. *The nation’s report card: Vocabulary results from the 2009 and 2011 NAEP reading assessments*. Washington, DC: Institute of Education, U. S. Department of Education; 2013. Retrieved from <http://nationsreportcard.gov/pdf/main2011/2013452.pdf>
- Nunnally, J.; Bernstein, I. *Psychometric theory*. 3. New York, NY: McGraw-Hill; 1994.
- O’Reilly T, Sabatini J, Bruce K, Pillarissetti S, McCormick C. Middle school reading assessment: Measuring what matters under a RTI Framework. *Reading Psychology*. 2012; 33:162–189. <http://dx.doi.org/10.1080/02702711.2012.631865>.
- Ouellette G, Beers A. A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*. 2010; 23:189–208. <http://dx.doi.org/10.1007/s11145-0089159-1>.
- Perfetti, C. Comprehending written language: A blueprint of the reader. In: Brown, C.; Hagoort, P., editors. *The neurocognition of language*. New York, NY: Oxford University Press; 1999. p. 167-208.
- Perfetti, C.; Landi, N.; Oakhill, J. The acquisition of reading comprehension skill. In: Snowling, M.; Hulme, C., editors. *The science of reading: A handbook*. New York, NY: Blackwell; 2005. p. 227-247. <http://dx.doi.org/10.1002/9780470757642.ch13>
- Perfetti C, Stafura J. Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*. 2014; 18:22–37. <http://dx.doi.org/10.1080/10888438.2013.827687>.
- Pinker S. Words and rules. *Lingua*. 1998; 106:219–242. [http://dx.doi.org/10.1016/S0024-3841\(98\)00035-7](http://dx.doi.org/10.1016/S0024-3841(98)00035-7).
- Protopapas A, Sideridis G, Mouzaki A, Simos P. Development of lexical mediation in the relation between reading comprehension and word reading skills in Greek. *Scientific Studies of Reading*. 2007; 11:165–197. <http://dx.doi.org/10.1080/10888430701344322>.
- Quinn, JM.; Wagner, RK.; Petscher, Y.; Lopez, D. Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development*. 2014. Advance online publication. <http://dx.doi.org/10.1111/cdev.12292>
- Raftery A. Bayesian model selection in social research. *Sociological Methodology*. 1995; 25:111–163. <http://dx.doi.org/10.2307/271063>.
- Rayner K, Foorman BR, Perfetti CA, Pesetsky D, Seidenberg MS. How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*. 2001; 2(Suppl):31–74. <http://dx.doi.org/10.1111/1529-1006.00004>. [PubMed: 26151366]
- Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*. 2007; 16(1, Suppl):19–31. <http://dx.doi.org/10.1007/s11136-007-9183-7>. [PubMed: 17479357]

- Rijmen F. Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*. 2010; 47:361–372. <http://dx.doi.org/10.1111/j.1745-3984.2010.00118.x>.
- Sabatini, J.; Bruce, K.; Steinberg, J. ETS Research Report No. RR-13–08. Princeton, NJ: ETS; 2013. SARA reading components tests, RISE form: Test design and technical adequacy.
- Schatschneider C, Fletcher J, Francis D, Carlson C, Foorman B. Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*. 2004; 96:265–282. <http://dx.doi.org/10.1037/0022-0663.96.2.265>.
- Semel, E.; Wiig, E.; Secord, W. The clinical evaluation of language fundamentals. 3. San Antonio, TX: Psychological Corp; 1995.
- Semel, E.; Wiig, E.; Secord, W. The clinical evaluation of language fundamentals, 4th ed.: Examiner's manual. San Antonio, TX: Pearson; 2003.
- Slobin, D.; Welsh, C. Elicited imitation as a research tool in developmental psycholinguistics. In: Ferguson, C.; Slobin, D., editors. *Studies of child language development*. New York, NY: Holt, Rinehart, & Winston; 1973. p. 485-497.
- Storch SA, Whitehurst GJ. Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*. 2002; 38:934–947. <http://dx.doi.org/10.1037/0012-1649.38.6.934>. [PubMed: 12428705]
- Tighe E, Schatschneider C. A dominance analysis approach to determining predictor importance in third, seventh, and tenth grade reading comprehension skills. *Reading and Writing*. 2014; 27:101–127. <http://dx.doi.org/10.1007/s11145-013-9435-6>.
- Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*. 1997; 40:1245–1260. <http://dx.doi.org/10.1044/jslhr.4006.1245>.
- Tomblin JB, Zhang X. The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*. 2006; 49:1193–1208. [http://dx.doi.org/10.1044/1092-4388\(2006\)086](http://dx.doi.org/10.1044/1092-4388(2006)086).
- Tunmer WE, Chapman JW. The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*. 2012; 45:453–466. <http://dx.doi.org/10.1177/0022219411432685>. [PubMed: 22293683]
- Vadasy P, Sanders E. Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology*. 2010; 102:786–803. <http://dx.doi.org/10.1037/a0019639>.
- van der Lely HK, Rosen S, McClelland A. Evidence for a grammar-specific deficit in children. *Current Biology*. 1998; 8:1253–1258. [http://dx.doi.org/10.1016/S0960-9822\(07\)00534-9](http://dx.doi.org/10.1016/S0960-9822(07)00534-9). [PubMed: 9822577]
- Verhoeven L, van Leeuwe J. Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*. 2008; 22:407–423. <http://dx.doi.org/10.1002/acp.1414>.
- Wagner, RK.; Herrera, SK.; Spencer, M.; Quinn, JM. Reconsidering the Simple View of Reading in an intriguing case of equivalent models: Commentary on Tunmer and Chapman (2012). *Journal of Learning Disabilities*. 2014. Advance online publication. <http://dx.doi.org/10.1177/0022219414544544>
- Wagner, R.; Torgesen, J.; Rashotte, C. Test of Word Reading Efficiency. 2. Austin, TX: Pro-Ed; 2012.
- Wiederholt, JL.; Bryant, BR. Gray Oral Reading Tests. 4. Austin, TX: ProEd; 2001.
- Williams, KT. Expressive Vocabulary Test. 2. Minneapolis, MN: Pearson Assessments; 2007. (EVT-2)
- Wolff U. Effects of a randomised reading intervention study: An application of structural equation modelling. *Dyslexia*. 2011; 17:295–311. <http://dx.doi.org/10.1002/dys.438>. [PubMed: 21953739]
- Woodcock, R. Woodcock Reading Mastery Test–Revised. Circle Pines, MN: American Guidance Services; 1987.
- Woodcock, R.; McGrew, K.; Mather, N. Woodcock-Johnson Tests of Achievement. 3. Itasca, IL: Riverside; 2001.

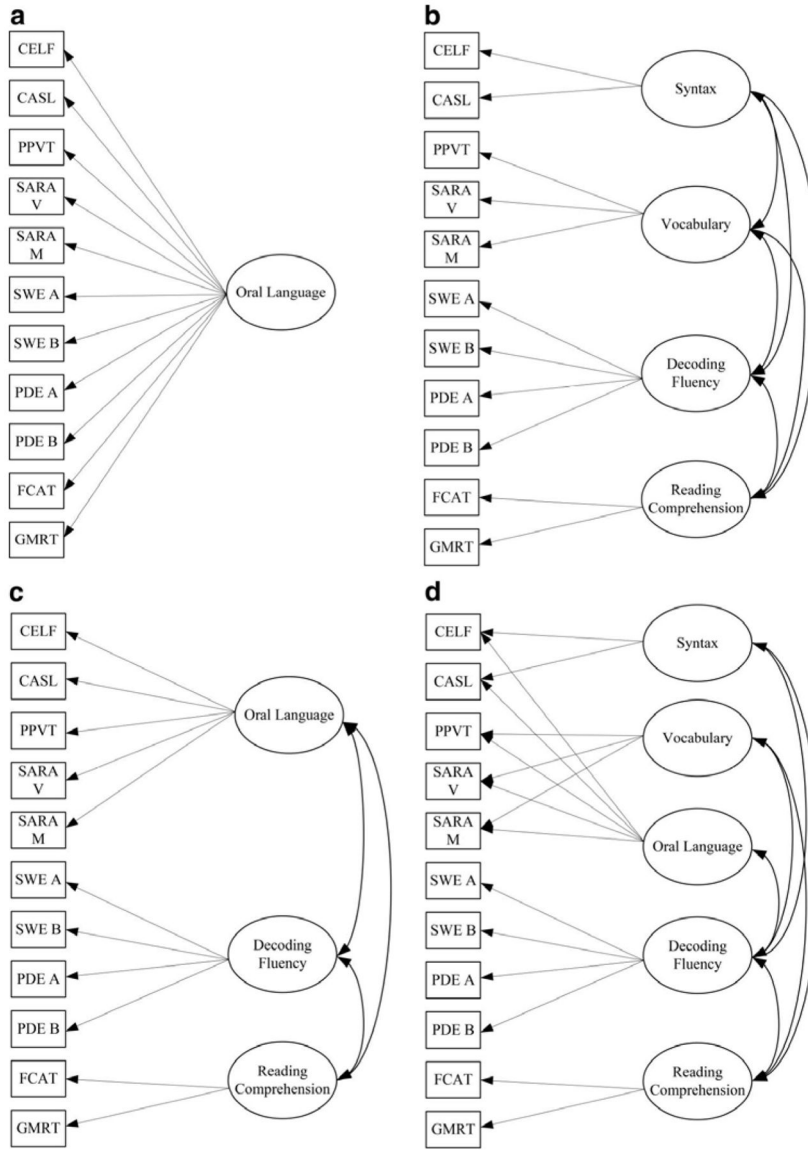
Woods C. Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*. 2009; 44:1–27. <http://dx.doi.org/10.1080/00273170802620121>.

Author Manuscript

Author Manuscript

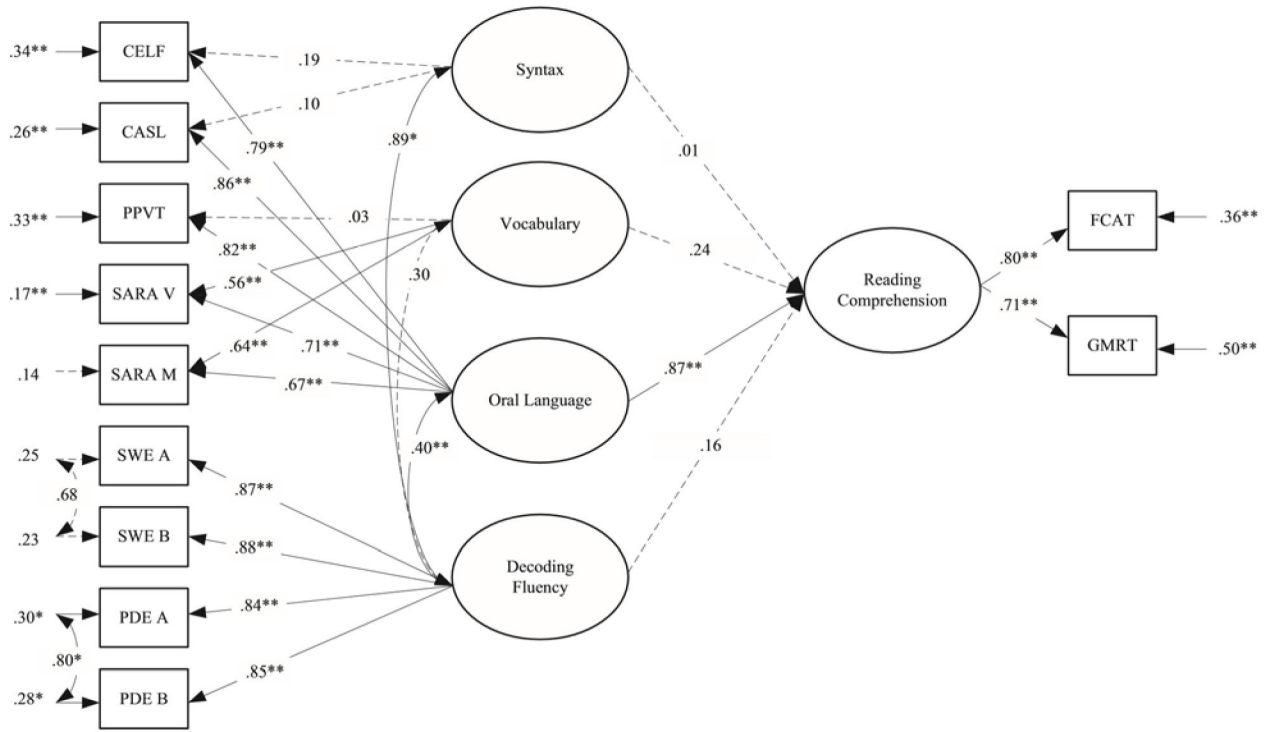
Author Manuscript

Author Manuscript



**Figure 1.** Figure 1a is Model 1, with one factor; Figure 1b is Model 2, with four correlated factors; Figure 1c is Model 3, with three correlated factors; Figure 1d is Model 4, the bifactor model. *Note.* CELF = Clinical Evaluation of Language Fundamentals, 4th ed. (Recalling Sentences subtest); CASL = Comprehensive Assessment of Spoken Language (Grammatical Judgment subtest); PPVT = Peabody Picture Vocabulary Test, 4th ed.; SARA = Study Aid and Reading Assessment (V = vocabulary; M = morphology); SWE = Sight Word Efficiency, Forms A and B; PDE = Phonetic Decoding Efficiency, Forms A and B (both from the Test of Word Reading Efficiency); FCAT = Florida Comprehensive Assessment Test; GMRT = Gates-MacGinitie Reading Test.





**Figure 2.** Grade 8 Structural Equation Model. *Note.* CELF = Clinical Evaluation of Language Fundamentals, 4th ed. (Recalling Sentences subtest); CASL = Comprehensive Assessment of Spoken Language (Grammatical Judgment subtest); PPVT = Peabody Picture Vocabulary Test, 4th ed.; SARA = Study Aid and Reading Assessment (V = vocabulary; M = morphology); SWE = Sight Word Efficiency, Forms A and B; PDE = Phonetic Decoding Efficiency, Forms A and B (both from the Test of Word Reading Efficiency); FCAT = Florida Comprehensive Assessment Test; GMRT = Gates-MacGinitie Reading Test. Significant paths are marked by solid lines and nonsignificant paths are marked by dashed lines.

**Table 1**

Final Number of Participants by Grade, Gender, and Race/Ethnicity

Grade	N	Female	Race/Ethnicity (in perspective)						White	Not reported
			Asian	Black	Hispanic	Indian	Multicultural	Other		
4	271	58	6	30	26	1	4	0	28	6
5	321	56	5	35	23	1	3	0	23	9
6	309	55	5	26	18	0	5	1	39	5
7	299	51	3	9	38	0	2	0	25	21
8	232	51	3	9	36	0	2	0	27	22
9	238	61	4	12	24	0	7	2	45	5
10	122	65	6	9	24	1	7	0	45	8

**Table 2**

Sample Size by Grade, Cohort, and Measure

Grade	Cohort	CELF RS	CASL GJT	PPVT	SARA Vocab	SARA Morph	TOWRE SWE Form A	TOWRE SWE Form B	TOWRE PDE Form A	TOWRE PDE Form B	FCAT	GMRT
4	3	177	177	180	179	179	176	176	175	175	258	128
5	2	213	210	207	205	205	210	210	210	210	274	270
6	1	100	0	101	94	94	96	96	95	95	61	92
3	3	134	139	138	138	138	139	139	139	139	202	92
7	2	199	196	198	195	194	199	199	199	199	257	254
8	3	151	158	158	153	153	151	151	151	151	220	97
9	1	101	0	101	95	95	101	101	101	101	67	90
2	2	90	80	84	86	85	90	90	90	90	123	126
10	3	74	81	84	78	78	70	70	70	70	117	57

**Note.** CELF = Clinical Evaluation of Language Fundamentals, 4th ed. (RS = Recalling Sentences); CASL = Comprehensive Assessment of Spoken Language (GJT = Grammatical Judgment subtest); PPVT = Peabody Picture Vocabulary Test, 4th ed.; SARA = Study Aid and Reading Assessment; TOWRE = Test of Word Reading Efficiency (SWE = Sight Word Efficiency, PDE = Phonetic Decoding Efficiency); FCAT = Florida Comprehensive Assessment Test; GMRT = Gates-MacGinitie Reading Test.

Table 3

## Correlations Between Measures

	1	2	3	4	5	6	7	8	9	10	11
1. CELF-RS	1.00										
2. CASL-GJT	.56	1.00									
3. PPVT	.48	.67	1.00								
4. SARA Vocabulary	.38	.51	.57	1.00							
5. SARA Morphology	.51	.48	.52	.71	1.00						
6. TOWRE SWE Form A	.50	.36	.35	.36	.54	1.00					
7. TOWRE SWE Form B	.50	.35	.34	.37	.55	.93	1.00				
8. TOWRE PDE Form A	.43	.33	.33	.41	.56	.77	.79	1.00			
9. TOWRE PDE Form B	.43	.35	.34	.42	.56	.76	.78	.94	1.00		
10. FCAT	.62	.51	.66	.60	.69	.58	.58	.54	.54	1.00	
11. GMRT	.55	.45	.54	.50	.65	.53	.55	.49	.49	.80	1.00

**Note.** All correlations are significant at the 0.01 level (2-tailed). CELF = Clinical Evaluation of Language Fundamentals, 4th ed. (RS = Recalling Sentences); CASL = Comprehensive Assessment of Spoken Language (GJT = Grammatical Judgment subtest); PPVT = Peabody Picture Vocabulary Test, 4th ed.; SARA = Study Aid and Reading Assessment; TOWRE = Test of Word Reading Efficiency (SWE = Sight Word Efficiency, PDE = Phonetic Decoding Efficiency); FCAT = Florida Comprehensive Assessment Test; GMRT = Gates-MacGinitie Reading Test.

**Table 4**  
 Number of Participants, Means, Standard Deviations, Minimum, and Maximum by Grade for All Measures

Grade	Statistic	CELF RS	CASL GJT	PPVT	SARA Vocab	SARA Morph	TOWRE SWE Form A	TOWRE SWE Form B	TOWRE PDE Form A	TOWRE PDE Form B	FCAT	GMRT
4	N	177	177	180	179	179	176	176	175	175	258	128
	Mean	60.24	95.67	99.44	28.25	20.13	68.85	68.74	36.39	37.28	217.09	497.56
	Std. Dev.	13.05	12.79	14.96	6.64	6.44	8.32	8.67	9.94	10.45	18.31	34.78
	Min.	30	53	57	8	6	36	35	9	4	154	417
	Max.	88	125	139	38	31	92	87	58	55	269	619
5	N	213	210	207	205	205	210	210	21	210	274	270
	Mean	61.38	91.50	99.16	30.02	21.24	71.60	71.72	38.58	39.10	224.30	509.51
	Std. Dev.	12.69	13.45	15.82	6.59	6.64	12.29	11.91	12.29	12.58	21.97	36.15
	Min.	23	40	45	7	1	11	7	3	5	161	411
	Max.	88	125	146	38	32	104	97	63	65	277	596
6	N	234	139	239	232	232	235	235	234	234	263	184
	Mean	68.37	90.94	102.16	27.58	22.13	76.86	76.40	41.89	42.41	227.05	510.85
	Std. Dev.	12.33	11.19	13.47	7.21	6.42	9.63	9.70	10.74	11.34	20.33	37.21
	Min.	36	72	67	0	7	42	46	13	10	167	363
	Max.	92	120	135	38	32	99	100	63	64	283	617
7	N	199	196	198	195	194	199	199	199	199	257	254
	Mean	66.32	89.46	96.41	28.92	22.62	75.78	74.87	41.54	41.57	230.40	528.35
	Std. Dev.	14.85	13.65	16.02	6.74	7.20	11.48	11.06	11.84	12.16	22.44	36.83
	Min.	22	40	37	9	5	31	34	12	8	171	443
	Max.	93	149	135	38	32	108	103	64	65	289	626
8	N	151	158	158	153	153	151	151	151	151	220	97
	Mean	68.70	87.78	98.63	31.00	24.84	78.19	77.95	44.53	44.64	236.30	539.46
	Std. Dev.	14.28	12.27	15.18	6.98	6.88	10.78	11.35	11.12	11.39	21.46	35.23
	Min.	22	43	43	8	8	48	45	12	9	178	463
	Max.	95	113	131	38	32	107	104	63	66	296	660
9	N	191	80	185	181	180	191	191	191	191	190	216
	Mean	74.61	92.94	97.90	25.56	25.12	83.02	82.50	46.22	46.56	240.31	548.54
	Std. Dev.	13.76	10.33	15.63	6.19	4.81	11.02	11.44	11.83	11.98	23.58	38.27
	Min.	22	43	43	8	8	48	45	12	9	178	463
	Max.	95	113	131	38	32	107	104	63	66	296	660

Grade	Statistic	CELF RS	CASL G-JT	PPVT	SARA Vocab	SARA Morph	TOWRE SWE Form A	TOWRE SWE Form B	TOWRE PDE Form A	TOWRE PDE Form B	FCAT	GMRT
10	Min.	37	56	51	8	8	47	47	9	5	178	363
	Max.	96	109	129	35	30	108	107	66	66	302	626
	N	74	81	84	78	78	70	70	70	70	117	57
	Mean	78.54	93.63	106.33	29.69	25.37	88.66	88.57	51.91	52.73	257.20	569.81
	Std. Dev.	11.71	10.89	13.74	4.28	5.06	10.58	9.01	9.47	10.32	18.38	21.75
	Min.	45	49	55	13	5	55	59	26	28	202	515
Max.	94	117	132	35	30	108	106	65	66	302	617	

**Note.** CELF = Clinical Evaluation of Language Fundamentals, 4th ed. (RS = Recalling Sentences); CASL = Comprehensive Assessment of Spoken Language (GJT = Grammatical Judgment subtest); PPVT = Peabody Picture Vocabulary Test, 4th ed.; SARA = Study Aid and Reading Assessment; TOWRE = Test of Word Reading Efficiency (SWE = Sight Word Efficiency, PDE = Phonetic Decoding Efficiency); FCAT = Florida Comprehensive Assessment Test; GMRT = Gates-MacGinitie Reading Test. Developmental scale scores were used for the FCAT and GMRT; in the absence of developmental scale scores, standard scores were used for the CASL and PPVT-4; due to the unavailability of developmental scale or standard scores, total correct raw scores were used for CELF-4, TOWRE-2, and SARA measures.

**Table 5**

Fit Indices for Models by Grade

Grades	Model	$\chi^2$	df	p	AIC	Adj BIC	RMSEA	LB	UB	CFI	TLI	SRMR
4	One factor	255.94	42	0.00	14094	14110	0.14	0.12	0.15	0.85	0.81	0.14
	Four factor	142.85	39	0.00	13987	14004	0.10	0.08	0.12	0.93	0.90	0.08
	Three factor	110.91	36	0.00	13961	13979	0.09	0.07	0.11	0.95	0.92	0.06
	Bifactor	51.07	30	0.01	13914	13934	0.05	0.03	0.07	0.99	0.97	0.03
5	One factor	236.20	42	0.00	17646	17667	0.12	0.11	0.14	0.91	0.88	0.08
	Four factor	119.41	39	0.00	17536	17558	0.08	0.06	0.10	0.96	0.95	0.06
	Three factor	99.79	36	0.00	17522	17547	0.07	0.06	0.09	0.97	0.96	0.05
	Bifactor	42.62	30	0.06	17477	17505	0.04	0.00	0.06	0.99	0.99	0.02
6	One factor	260.22	42	0.00	17386	17405	0.13	0.12	0.15	0.89	0.86	0.11
	Four factor	82.32	39	0.00	17214	17235	0.06	0.04	0.08	0.98	0.97	0.04
	Three factor	75.30	36	0.00	17213	17236	0.06	0.04	0.08	0.98	0.97	0.04
	Bifactor	39.49	30	0.12	17189	17215	0.03	0.00	0.06	1.00	0.99	0.03
7	One factor	211.90	42	0.00	16821	16840	0.12	0.10	0.13	0.91	0.89	0.10
	Four factor	94.84	39	0.00	16710	16730	0.07	0.05	0.09	0.97	0.96	0.06
	Three factor	83.81	36	0.00	16705	16727	0.07	0.05	0.09	0.98	0.96	0.05
	Bifactor	34.22	30	0.27	16667	16692	0.02	0.00	0.05	1.00	1.00	0.03
8	One factor	201.26	42	0.00	12161	12171	0.13	0.11	0.15	0.89	0.86	0.12
	Four factor	107.81	39	0.00	12073	12084	0.09	0.07	0.11	0.95	0.93	0.06
	Three factor	72.65	36	0.00	12044	12056	0.07	0.04	0.09	0.98	0.96	0.05
	Bifactor	22.29	30	0.84	12006	12019	0.00	0.00	0.03	1.00	1.01	0.03
9	One factor	233.79	42	0.00	14049	14060	0.14	0.12	0.16	0.90	0.87	0.12
	Four factor	91.33	39	0.00	13913	13925	0.08	0.06	0.10	0.97	0.96	0.08
	Three factor	86.96	36	0.00	13915	13927	0.08	0.06	0.10	0.97	0.96	0.07
	Bifactor	53.92	30	0.00	13894	13908	0.06	0.03	0.08	0.99	0.98	0.05
10	One factor	88.95	42	0.00	5814	5801	0.10	0.07	0.12	0.93	0.91	0.18
	Four factor	58.05	39	0.03	5788	5774	0.06	0.02	0.09	0.97	0.96	0.13
	Three factor	40.46	36	0.28	5777	5763	0.03	0.00	0.07	0.99	0.99	0.11
	Bifactor	23.97	31	0.81	5771	5754	0.00	0.00	0.04	1.00	1.02	0.13

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Note.** AIC = Akaike information criterion; BIC = Bayesian information criterion; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual. The bifactor measurement model specifications varied by grade based on significance testing ( $p = .05$ ).



**Table 6**

**Structural Equation Model Results by Grade**

Grade	Outcome	Predictors	Estimate	SE	Est./SE	Standardized	p-value
4	Reading comprehension	Syntax	3.45	0.49	7.11	0.50	0.00
		Vocabulary	0.85	0.68	1.25	0.12	0.21
	Decoding fluency	Oral language	5.13	0.75	6.84	0.75	0.00
		Syntax	1.46	0.87	1.68	0.21	0.09
5	Reading comprehension	Syntax	-0.24	0.18	-1.40	-0.24	0.16
		Vocabulary	0.43	0.08	5.70	0.43	0.00
	Decoding fluency	Oral language	0.44	0.11	4.22	0.44	0.00
		Syntax	0.36	0.53	0.68	0.16	0.50
6	Reading comprehension	Vocabulary	0.38	0.53	0.72	0.17	0.47
		Oral language	1.76	0.57	3.06	0.76	0.00
	Decoding fluency	Syntax	0.31	0.72	0.42	0.13	0.67
		Vocabulary	0.33	0.11	3.07	0.33	0.00
7	Reading comprehension	Vocabulary	0.43	0.12	3.50	0.43	0.00
		Oral language	0.63	0.08	7.68	0.63	0.00
	Decoding fluency	Syntax	-0.19	0.14	-1.36	-0.10	0.17
		Vocabulary	0.15	0.17	0.86	0.08	0.39
7	Reading comprehension	Oral language	1.37	0.26	5.29	0.72	0.00
		Syntax	0.32	0.19	1.62	0.17	0.10
	Decoding fluency	Syntax	-0.06	0.10	-0.63	-0.06	0.53
		Vocabulary	0.29	0.12	2.52	0.29	0.01
7	Reading comprehension	Oral language	0.59	0.07	8.57	0.59	0.00
		Syntax	-0.24	0.22	-1.11	-0.07	0.27
	Decoding fluency	Vocabulary	0.62	0.30	2.09	0.18	0.04
		Oral language	3.20	0.43	7.44	0.92	0.00
7	Reading comprehension	Decoding fluency	0.12	0.26	0.45	0.03	0.65
		Syntax	-0.05	0.09	-0.55	-0.05	0.59
	Decoding fluency	Vocabulary	0.13	0.10	1.35	0.13	0.18
		Oral language	0.66	0.06	11.36	0.66	0.00

Grade	Outcome	Predictors	Estimate	SE	Est./SE	Standardized	p-value
8	Reading comprehension	Syntax	0.10	1.20	0.08	0.01	0.94
		Vocabulary	1.83	1.69	1.08	0.24	0.28
	R <sup>2</sup> = .983	Oral language	6.65	1.93	3.44	0.87	0.00
		Decoding fluency	1.21	1.65	0.74	0.16	0.46
		Syntax	0.89	0.45	1.98	0.89	0.05
		Vocabulary	0.30	0.20	1.49	0.30	0.14
9	Reading comprehension	Oral language	0.40	0.14	2.85	0.40	0.00
		Syntax	-2.40	1.42	-1.69	-0.16	0.09
	R <sup>2</sup> = .995	Vocabulary	-3.45	1.85	-1.87	-0.24	0.06
		Oral language	12.71	1.65	7.72	0.87	0.00
		Decoding fluency	2.49	1.66	1.51	0.17	0.13
		Syntax	0.26	0.15	1.70	0.26	0.09
10	Reading comprehension	Vocabulary	0.32	0.10	3.14	0.32	0.00
		Oral language	0.58	0.07	8.77	0.58	0.00
	R <sup>2</sup> = .960	Syntax	-0.90	1.20	-0.74	-0.18	0.46
		Vocabulary	1.09	1.59	0.69	0.22	0.49
		Oral language	4.82	1.84	2.62	0.96	0.01
		Decoding fluency	-0.16	1.55	-0.10	-0.03	0.92
Decoding fluency	Syntax	-0.02	0.28	-0.08	-0.02	0.93	
	Vocabulary	0.10	0.28	0.37	0.10	0.71	
		Oral language	0.60	0.13	4.45	0.60	0.00

**Note.** Nonsignificant residual variances were constrained to 0 (Comprehensive Assessment of Spoken Language (CASL) in grades 7, 9, and 10; Study Aid and Reading Assessment (SARA) Morphology in grade 4; SARA Vocabulary in grade 7) to improve model fit. A nonsignificant error covariance between Test of Word Reading Efficiency (TOWRE) Sight Word Efficiency Form A and B in grade 7 was also removed from the model specification to improve fit.