Published in final edited form as:

Read Writ. 2015 April 1; 28(4): 439–466. doi:10.1007/s11145-014-9532-1.

Applying a Multiple Group Causal Indicator Modeling Framework to the Reading Comprehension Skills of Third, Seventh, and Tenth Grade Students

Elizabeth L. Tighe, **Richard K. Wagner**, and **Christopher Schatschneider**

Florida State University

Abstract

This study demonstrates the utility of applying a causal indicator modeling framework to investigate important predictors of reading comprehension in third, seventh, and tenth grade students. The results indicated that a 4-factor multiple indicator multiple indicator cause (MIMIC) model of reading comprehension provided adequate fit at each grade level. This model included latent predictor constructs of decoding, verbal reasoning, nonverbal reasoning, and working memory and accounted for a large portion of the reading comprehension variance (73% to 87%) across grade levels. Verbal reasoning contributed the most unique variance to reading comprehension at all grade levels. In addition, we fit a multiple group 4-factor MIMIC model to investigate the relative stability (or variability) of the predictor contributions to reading comprehension across development (i.e., grade levels). The results revealed that the contributions of verbal reasoning, nonverbal reasoning, and working memory to reading comprehension were stable across the three grade levels. Decoding was the only predictor that could not be constrained to be equal across grade levels. The contribution of decoding skills to reading comprehension was higher in third grade and then remained relatively stable between seventh and tenth grade. These findings illustrate the feasibility of using MIMIC models to explain individual differences in reading comprehension across the development of reading skills.

Keywords

causal indicator modeling; decoding; reading comprehension; nonverbal reasoning; verbal reasoning; working memory

---

Comprehension, the overall goal in reading, requires readers to acquire information, synthesize and integrate text, and go beyond recognizing individual words to actively construct and extract meaning from text. Therefore, reading comprehension is a complex process, which involves many higher and lower order cognitive skills (for a review see Cain & Oakhill, 2007). Past research has identified numerous cognitive predictors of reading comprehension at various points in development (grade levels): listening comprehension (Sticht & James, 1984), decoding (Chen & Vellutino, 1997), fluency (Tighe & Schatschneider, 2014), phonological skills (de Jong & van der Leij, 2002), reasoning

---

Correspondence should be addressed to Elizabeth Tighe, Department of Psychology, Florida State University, 1107 W. Call St, Tallahassee, FL, 32306-4301, USA. tighe@psy.fsu.edu.

(Schatschneider, Harrell, & Buck, 2007), vocabulary knowledge (Beck & McKeown, 1991), morphological awareness (Deacon & Kirby, 2004), working memory (Cain, Oakhill, & Bryant, 2004), and comprehension monitoring and inferencing (Oakhill & Cain, 2012). However, fewer studies have examined important reading-related predictors and the stability (or variability) of the magnitudes and contributions of these predictors to reading comprehension across development, particularly in middle and high school students. The purpose of the current study was twofold. First, we introduce the technique of causal indicator modeling and demonstrate the utility of fitting separate causal indicator models by grade level (third, seventh, and tenth) to investigate the magnitudes of the joint and unique variance estimates of decoding, verbal reasoning, nonverbal reasoning, and working memory to reading comprehension. Second, we fit a multiple group causal indicator model to assess the relative stability (or variability) of the predictor contributions to reading comprehension across the three distinct developmental grades.

Key Contributors to Individual Differences in Reading Comprehension

The first goal of the current study was to determine important predictors of reading comprehension that have been identified across multiple grade levels to include in our causal indicator models. We review the extant literature on the contributions of decoding, language skills (listening comprehension and verbal reasoning), nonverbal reasoning skills, and working memory below. From a theoretical and empirical standpoint, it is important to identify plausible predictors of reading comprehension at each grade level in order to properly specify our causal indicator models. In addition, we provide evidence of how the contributions of these predictors may vary across different grade levels because the second goal of our study was to test for measurement invariance in our causal indicator models to assess how stable or variable the predictor contributions to reading comprehension were across our grade levels.

Decoding and linguistic comprehension skills

Several studies have reported that lower-order, word reading skills and higher-order, language skills are important core processes in the development of reading comprehension (Goff, Pratt, & Ong, 2005; Tilstra, McMaster, Van den Broek, Kendeou, & Rapp, 2009; Vellutino, Tunmer, Jaccard, & Chen, 2007; Verhoeven & van Leeuwe, 2008). The Simple View of Reading (SVR), postulates that reading comprehension represents the product of an individual's decoding and linguistic comprehension skills ($R = D \times C$) (Gough & Tunmer, 1986; Hoover & Gough, 1990). Past research has found that the SVR component skills account for roughly 40% to 80% of the variance in reading comprehension skills across multiple grade levels and multiple measures of word reading and language skills (Catts, Hogan, & Adlof, 2005; Dreyer & Katz, 1992; Tilstra et al., 2009).

Beyond identifying word reading and language skills as important component skills of reading comprehension, several studies have explored if the contributions of these skills vary as a function of development (i.e., grade level) (Cutting & Scarborough, 2006; de Jong & van der Leij, 2002; Goff et al., 2005; Verhoeven & van Leeuwe, 2008). There is considerable consensus that the relationship between decoding and reading comprehension is stronger in the early elementary school years; whereas the relationship between linguistic comprehension and reading comprehension begins to increase around third grade once decoding skills have been mastered (Adlof, Catts, & Little, 2006; Diakidoy, Stylianou, Karefillidou, & Papageorgiou, 2005; Kershaw & Schatschneider, 2012; Tilstra et al., 2009; Verhoeven & van Leeuwe, 2008). For example, Tilstra et al. (2009) examined the predictive utility of decoding, listening comprehension, fluency, and verbal proficiency (vocabulary) to reading comprehension in fourth, seventh, and ninth graders. The contribution of decoding skills to reading comprehension decreased between fourth grade and the upper grade levels (seventh and ninth); whereas the contribution of listening comprehension to reading comprehension increased between fourth and seventh grade and remained stabled between seventh and ninth grade. Similarly, Adlof et al. (2006) reported a decrease in the importance of decoding skills to reading comprehension and an increase in the importance of listening comprehension skills to reading comprehension between the second and eighth grade levels. In sum, the past research indicates developmental changes in the contributions of the SVR component skills to reading comprehension.

Verbal and nonverbal reasoning

Higher-order reasoning skills involve the ability to generate inferences from complex text and typically develop as children progress through middle school (Cain & Oakhill, 1998). Verbal reasoning skills align with the linguistic comprehension component of the SVR and typically encompass listening comprehension, vocabulary knowledge, and verbal IQ skills. Verbal reasoning skills are also closely associated with specific discourse skills, such as inference and integration, comprehension monitoring, and knowledge of text structure (Oakhill & Cain, 2012). Nonverbal reasoning skills (also referred to as performance IQ) assess inferencing and problem-solving abilities through spatial skills, abstract concepts, and visuomotor coordination (Kershaw & Schatschneider, 2012). Research has consistently identified verbal reasoning (Jensen, 1980; Oakhill & Cain, 2012; Oakhill, Cain, & Bryant, 2003; Schatschneider et al., 2007; Sternberg & Powell, 1983; Tighe & Schatschneider, 2014) and nonverbal reasoning (Adlof, Catts, & Lee, 2010; Asbell, Donders, Van Tubbergen, & Warschausky, 2010; Fuchs et al., 2012; Kershaw & Schatschneider, 2012; Tighe & Schatschneider, 2014; Tiu, Thompson, & Lewis, 2003) as important predictors of reading comprehension across multiple grade levels. Further, verbal reasoning has been identified as a unique predictor of reading comprehension after controlling for fluency (Tighe & Schatschneider, 2014), working memory (Alloway & Alloway, 2010; Tighe & Schatschneider, 2014), and nonverbal reasoning (Alloway & Alloway, 2010; Oakhill et al., 2003; Tighe & Schatschneider, 2014). Likewise, nonverbal reasoning has emerged as a unique predictor of reading comprehension independent of decoding (Kershaw & Schatschneider, 2012; Tiu et al., 2003), linguistic comprehension (Kershaw & Schatschneider, 2012; Tiu et al., 2003), and processing speed (Tiu et al., 2003).

Moreover, the contributions of verbal and nonverbal reasoning to reading comprehension may vary dependent upon grade level (Adlof et al., 2010; Schatschneider et al., 2007; Tighe & Schatschneider, 2014). For example, Tighe and Schatschneider (2014) reported a shift in the most important predictors of reading comprehension across third, seventh, and tenth grade students. Fluency contributed the most unique variance (25%) to third grade reading comprehension; however, reasoning (comprised of verbal and nonverbal reasoning) was the sole most important predictor of tenth grade reading comprehension (30% unique variance). Similarly, Adlof et al. (2010) found that nonverbal reasoning measures assessed at kindergarten were stronger predictors of eighth grade reading comprehension as compared to second grade reading comprehension skills. Taken together, these findings indicate a possible developmental shift in the importance of higher-order, reasoning skills, such that reasoning skills become increasingly important to reading comprehension during the middle and high school years.

Working memory

Working memory is another construct that has been consistently included as a predictor of reading comprehension (correlations ranging from .3–.9) across several grade levels (Cain et al., 2004; Molloy, 1997; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000; Oakhill & Cain, 2012; Swanson & Howell, 2001). Working memory is a limited capacity store that maintains, manipulates, and processes currently held information (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980). The Working Memory Resource Hypothesis suggests that strong working memory capacity underlies the development of successful reading comprehension abilities (Swanson & O'Connor, 2009). Readers need to be able to actively store and process concurrent text information in order to activate background knowledge, generate inferences, and engage in comprehension monitoring during reading (Cain, Oakhill, Bryant, 2004; Oakhill & Cain, 2012; Swanson & O'Connor, 2009).

Research on the Working Memory Resource Hypothesis has produced mixed findings. For example, several studies have reported that working memory exerts a direct influence on reading comprehension after controlling for word reading skills, vocabulary, and verbal skills (Cain et al., 2004; Seigneuric & Ehrlich, 2005; Seigneuric et al., 2000; Swanson & Berninger, 1995). Further, working memory explains additional reading comprehension variance independent of nonverbal IQ (Alloway & Alloway, 2010) and discourse skills (integration and inference, comprehension monitoring, knowledge of text structure) (Cain et al., 2004). Other studies have found that working memory exerts an indirect influence on reading comprehension through the mediators of fluency (Swanson & O'Connor, 2009), verbal and semantic skills (Nation, Adams, Bowyer-Crane, & Snowling, 1999), and attentional control (McVay & Kane, 2012). However, some research has suggested that working memory does not contribute uniquely to reading comprehension after controlling for additional component skills: decoding, oral language skills, reading speed, reasoning, fluency, and rapid serial naming (Cutting & Scarborough, 2006; Goff et al., 2005; Tighe & Schatschneider, 2014).

The diverse findings regarding working memory as a predictor of reading comprehension make it difficult to evaluate the relationship between these two skills and to assess if this relationship changes as a function of development. One longitudinal study examined the contribution of working memory to reading comprehension across development (first through third grade) after controlling for decoding and vocabulary knowledge (Seigneuric & Ehrlich, 2005). Working memory did not contribute unique variance to first and second grade reading comprehension; however, working memory accounted for an additional 6% of the variance in third grade reading comprehension. These researchers concluded that once decoding skills are mastered in the earlier grades, higher-order processes such as vocabulary knowledge and working memory become increasingly important as a function of grade level. This finding supports a developmental shift in the importance of working memory to reading comprehension; however, more research is needed across multiple grade levels.

Causal Indicator Models

The present study explored the utility of building causal indicator models of reading comprehension in third, seventh, and tenth grade students. Causal indicator modeling allowed us to assess the magnitudes of the joint and unique contributions of four latent constructs to reading comprehension: decoding, verbal reasoning, nonverbal reasoning, and working memory. In the traditional, confirmatory factor analysis (CFA) approach, effect indicators reflect the variance in the latent construct. For example, the variance in the latent construct (reading comprehension) explains individual differences in the observed indicators (measures of decoding, verbal reasoning, nonverbal reasoning, or working memory) (Kline, 2011). In contrast, causal indicator models reverse causal flow and allow the indicators to cause the variability in the latent construct. For example, causal indicators (measures of decoding, verbal reasoning, nonverbal reasoning or working memory) comprise the variability in the latent construct (reading comprehension). However, the causal indicators do not account for all of the variability because the latent construct still includes a disturbance term for residual variance (Bollen & Bauldry, 2011; Bollen & Davis, 2009).

We utilized the three conceptual checks outlined by Bollen and Bauldry (2011) to determine whether to use a traditional CFA approach or a causal indicator modeling approach. First, we needed to determine if a change in the latent construct corresponds to a change in all of the observed indicators (traditional model) or if a change in an indicator corresponds to a change in the latent construct (causal indicator model). This cannot be accomplished by statistical testing; however, this can be determined based on theory and hypotheses. In our study, we assessed the underlying nature of reading comprehension by determining important component skills of reading comprehension. Our literature review indicated that all four of our constructs have been identified as important predictors of reading comprehension across multiple grade levels. Thus, we assumed that a change in any of the observed measures of decoding, verbal reasoning, nonverbal reasoning, and working memory would result in a change in the amount of variance accounted for in reading comprehension.

For the second conceptual check, we needed to evaluate the essentialness and interchangeableness of the indicators (i.e., does removing any of the indicators alter the variance accounted for in the latent construct?). In our models, we assumed that the four causal indicators work jointly. Therefore, we expected that dropping any of the indicators from the model would result in a change in the correlations between the remaining indicators as well as a change in the reading comprehension variance accounted for by the remaining indicators. Thus, our four causal indicators are all essential to comprising reading comprehension and the indicators are not interchangeable.

The final conceptual check states that the indicators do not necessarily need to be correlated with each other in a causal indicator modeling approach. In our models, we anticipated that the component skills would be correlated with each other; however, we hypothesized that these skills would be correlated to varying degrees. For example, verbal reasoning may be more highly correlated with nonverbal reasoning than the correlation between working memory and decoding. Based on these three conceptual checks, we determined that utilizing a causal indicator modeling approach was preferred for our hypotheses and research questions.

In addition to the three theoretically driven conceptual checks, we utilized a two-step empirical check proposed by Treiblmaier, Bentler, and Mair (2011) to assess the appropriateness of applying a causal indicator modeling approach. This empirical check relies on canonical correlation analysis (CCA) for proper model identification and to evaluate that the causal indicators (e.g., our predictors of decoding, verbal reasoning, nonverbal reasoning, and working memory) are accounting for the variability in the latent construct (reading comprehension). The first step specifies that the researcher form two sets of maximally correlated linear composites from the observed indicators of the predictor constructs and calculate the canonical correlation between the composites. The second step indicates that the researcher should correlate the composites obtained in step one with factor scores of the latent construct (reading comprehension). The higher the correlation between the composites and the latent variable, the more justified the researcher is in concluding that the causal indicators are appropriately specified and accounting for the variance in the latent construct.

Causal indicator modeling has recently been explored in an educational context to build reading comprehension models (Tighe, 2012; Wagner, 2013). Tighe (2012) used causal indicator models to investigate the contributions of morphological awareness and vocabulary knowledge to reading comprehension in a sample of Adult Basic Education students. A 3-factor (real word morphology, pseudoword morphology, and vocabulary knowledge) model provided adequate fit to the data and these factors jointly accounted for 79% of the reading comprehension variance. Wagner (2013) successfully fit causal indicator reading comprehension models to nine different datasets of children (ranging from first through eighth grade). Across these datasets, several component skills (i.e., decoding, working memory, and oral language skills) were included as predictors of reading comprehensions; and these predictors jointly accounted for between 53% and 90% of the reading comprehension variance. The current study wanted to extend the previous literature on causal indicator modeling by: a. assessing 4-factor models at three distinct grade levels; and b. fitting a multiple group causal indicator model to assess the stability (or variability) of the four predictors across development.

Research Questions

The current study used causal indicator modeling to address three research questions:

1. How well does a 4-factor causal indicator model of reading comprehension that includes decoding, verbal reasoning, nonverbal reasoning, and working memory account for individual differences in reading comprehension?

2. What are the magnitudes of the joint and unique estimates of the causal indicators to reading comprehension?

3. Does the adequacy of a causal indictor model of reading comprehension and the contributions of the causal indicators vary by level of reading development (grade level)?

To address the first two research questions, we built 4-factor causal indicator models of reading comprehension separately by grade level. These models allowed us to assess the utility of fitting causal indicator models by determining the magnitudes of the joint and unique estimates of our four predictors to reading comprehension at each grade level. To address the third research question, we utilized a multiple group 4-factor causal indicator model to investigate measurement invariance and predictor stability across grade levels. This model allowed us to assess the stability of the observed measures of our four predictors as well as the stability of the contributions of the four predictors to reading comprehension across the three grade levels.

Method

Participants

The participants included 215 third graders, 188 seventh graders, and 182 tenth graders (total $N = 585$) attending 19 schools in three Florida educational districts during the 2002/2003 academic school year. Averaged across all grades, the sample consisted of approximately 54% female. The participants represented a range of ethnic backgrounds: 41% Caucasian, 38% African American, 17% Hispanic, 2% Asian, and 2% other/not specified. Approximately 36% of the sample qualified for free or reduced price lunch (a proxy for socioeconomic status). Participants were recruited for the study across 54 classrooms by parental consent forms that were sent home through classroom teachers.

Measures

Multiple measures of each construct of interest were obtained to permit analyses based on latent variables as opposed to observed variables.

Reading comprehension—Two group-administered measures of reading comprehension were available:

The Stanford Achievement Test – Ninth Edition (SAT-9) is a norm-referenced, untimed measure of reading comprehension (Harcourt Brace, 1997). Students were presented with literary, informational, and functional literacy passages followed by multiple-choice questions. The students answered a total of 54 multiple-choice questions, which encompassed passage understanding, interpretation, critical analysis, and awareness and usage of reading strategies. Scores are reported on a scale of 527 to 817. The reliability estimate for the SAT-9 is reported at .87.

The Sunshine State Standards Reading Comprehension subtest of the Florida Comprehensive Assessment Test (FCAT-SSS) is a norm-referenced measure, which includes six to eight informational and literary reading passages. Students were asked to read through the passages and answered between six to 11 multiple-choice questions per passage. These questions assessed words and phrases in context, the main idea of the passage, comparison and cause/effect relationships, and the ability to locate and evaluate information from multiple sources. Scores on this measure range from 100 to 500. The internal reliabilities for the reading subtest are .89, .90, and .85 for third, seventh, and tenth grade, respectively (Florida Department of Education, 2006).

Decoding—Two subtests of the Test of Word Reading Efficiency (TOWRE) were administered to assess decoding: the Phonemic Decoding Efficiency (PDE) subtest and the Sight Word Efficiency (SWE) subtest (Torgesen, Wagner, & Rashotte, 1999). The TOWRE is a norm-referenced, individually administered measure designed to assess word reading accuracy and fluency. The PDE is a timed subtest, in which the participants were presented with a list of pseudo-words. Participants were prompted to read aloud as many pseudo-words as possible in 45 seconds. The SWE is a timed subtest, in which the participants were presented with a list of real words. Participants were asked to read aloud as many real words as possible in 45 seconds. Test-retest reliability is reported at .90 for the PDE subtest and .97 for the SWE subtest.

Working memory—An adapted version of the Competing Language Processing Task, which included a reading span measure and a listening span measure, was used to assess working memory (Gaulin & Campbell, 1994). The reading span measure asked participants to read groups of three-word sentences and provide true or false responses. Participants were then prompted to recall the last word at the end of each sentence. For example, a student was given the following two sentences: “Candy is sweet. Triangles are round.” The participant would respond true or false to each sentence and then recall the final words: “sweet” and “round.” The groups of sentences increased in complexity, ranging from two to six sentences per group, as the task proceeded. Each task included a total of 42 items across the groups of sentences. Testing was stopped when fewer than half of the final words were recalled. The listening span measure was identical in format to the reading span measure, except that each sentence was read aloud to the participant. The alternate forms reliabilities for the reading span and listening span tasks were .66, .74 for third grade, .61, .79 for seventh grade, and .61, .70 for tenth grade.

Verbal reasoning—We included a composite score of three FCAT listening comprehension passages and two norm-referenced subtests from the Wechsler Abbreviated Scale of Intelligence (WASI) as observed indicators on the verbal reasoning factor in our models.

Listening Comprehension: Three FCAT passages were utilized to assess listening comprehension. The passages were read aloud to the participants and were shortened in length (not exceeding two minutes in total read time). After each passage, the examiner read aloud a series of multiple-choice questions and the participants marked down their answers on a scoresheet. The participants answered a total of 12 multiple-choice questions across the three passages. Reliability coefficients were .88 in third grade, .82 in seventh grade, and .85 in tenth grade.

Wechsler Abbreviated Scale of Intelligence (WASI): Two subtests of the WASI were used to assess verbal reasoning: vocabulary and similarities (Psychological Corporation, 1999). The first four items of the vocabulary subtest were low-end picture items, which required the participant to name pictures that were displayed one at time. The remaining 38 items were words presented aloud and visually that the participant needed to define aloud. Reliability coefficients ranged from .88 for third grade, .86 for seventh grade, and .83 for tenth grade. The similarities subtest measures abstract verbal reasoning abilities. Participants were asked to identify relationships between pairs of words that were presented either verbally or with pictures. Reliability coefficients were .89, .85, and .83 for third, seventh, and tenth grade, respectively.

Nonverbal reasoning—Two subtests of the WASI were used to assess nonverbal reasoning: matrix reasoning and block design (Psychological Corporation, 1999). The matrix reasoning subtest measures nonverbal fluid reasoning and general intellectual abilities. This task presented participants with a series of 35 incomplete patterns, each with five answer choices. Participants completed the task by either pointing to or stating the number of their answer choice. Reliability coefficients were .93 for third grade, .89 for seventh grade, and .86 for tenth grade. The block design subtest is a measure of perceptual organization and general intelligence that is designed to tap abilities related to spatial visualization, visuomotor coordination, and abstract conceptualization. Participants were given a set of blocks, which included blocks that had all white sides, blocks that had all red sides, and blocks that had red and white sides. The participants were asked to arrange the blocks in accordance with a pattern shown to them on a card by the examiner. Reliability coefficients were .92 for third grade, .92 for seventh grade, and .89 for tenth grade.

Procedure

The two-hour battery of measures was individually administered to students after they completed the FCAT in the spring of 2003. All of the measures, with the exception of listening comprehension and working memory, were norm-referenced assessments. The order of the administration of the assessments was counterbalanced. All testers underwent rigorous training and reached an acceptable level of test administration proficiency prior to assessing participants.

Results

Checking for Data Issues and Descriptive Statistics

The dataset was examined for missing data, outliers, skewness, and kurtosis. Univariate outliers were identified and brought to the boundaries of the median +/− two interquartile ranges. Across the three grades and 11 variables per grade, there were a total of 46 univariate outliers that were adjusted. No bivariate outliers were detected in the examination of the scatterplots of all pairs of variables. All skewness and kurtosis values fell within an acceptable range (+/− 2), indicating that all variables were normally distributed. There were 31 missing values in third grade, 20 missing values in seventh grade, and 9 missing values in tenth grade. Because there were relatively few missing data points and no more than 8 missing data points in any single variable, maximum likelihood (ML) estimation was used.

Means and standard deviations for each measure by grade are listed in Table 1. Correlations among the measures for each grade are presented in Tables 2, 3, and 4. The correlations show relatively high, positive relationships among the component reading skill measures across all grades. Additionally, all component skill measures exhibit moderate to high correlations with the measures of reading comprehension.

MIMIC Models of Reading Comprehension

To address our first two research questions, we fit 4-factor multiple indicator multiple indicator cause (MIMIC) models of reading comprehension separately by grade utilizing Version 6.12 of the Mplus statistical package (Muthén & Muthén, 1998–2010). The 4-factor MIMIC models included causal indicators of decoding, verbal reasoning, nonverbal reasoning, and working memory. MIMIC models are the preferred version of causal indicator models because they allow for proper model identification. A MIMIC model specifies that a model will be identified if the latent construct (reading comprehension) emits at least two arrows to reliable observed indicators. To resolve this identification problem, two observed indicators (SAT-9 and FCAT measures) were loaded onto the reading comprehension latent variable for each grade level (Bollen & Bauldry, 2011; Treiblmaier et al., 2011). Scale dependency was handled by fixing one indicator per latent variable to 1.0. We relied on the Hu and Bentler (1998) standards to evaluate model fit: Tucker Lewis Index (TLI) and Comparative Fit Index (CFI) values greater than .95, Root Mean Square Error of Approximation (RMSEA) values less than .08, and Standardized Root Mean Square Residual (SRMR) values of less than .05. We report model fit statistics, total R^2 values, and the magnitudes and significance values of the unique and independent estimates of the causal indicators by grade below.

Third grade—The 4-factor MIMIC model of reading comprehension provided adequate fit to the data as evidenced by the model fit indices ($\chi^2(34) = 69.94$, $p < .001$, CFI = .974, TLI = .958, RMSEA = .070, and SRMR = .030). All factor loading were significant (*p*s < .001) and all were above .63. Jointly, the four causal indicators accounted for approximately 77% of the variance in reading comprehension (Figure 1). Decoding and verbal reasoning contributed additional unique variance (5.7% and 9.5%, respectively) after controlling for the other predictors in the model. Looking at Figure 1, it appears as though nonverbal reasoning and working memory are not significant predictors of reading comprehension (β =. 144, $p = .066$; β =.077, $p = .285$, respectively). In conjunction with all other predictors included in the model, nonverbal reasoning and working memory do not account for any unique variance in predicting reading comprehension. However, in separate models without the other predictors included, each causal indicator contributes significant variance to reading comprehension: decoding (β =. 744, $p< .001$, $R^2 = .561$), verbal reasoning ($\beta = .816$, $p< .001$, $R^2 = .665$), nonverbal reasoning ($\beta = .662$, $p< .001$, $R^2 = .438$), and working memory ($\beta = .547$, $p< .001$ $R^2 = .299$). Overall, verbal reasoning emerged as the most important predictor of third grade reading comprehension; however, all of the causal indicators were significantly associated with reading comprehension.

Seventh grade—The 4-factor MIMIC model also yielded adequate fit in our seventh grade sample ($\chi^2(34) = 70.27$, $p < .001$, CFI = .965 TLI = .943, RMSEA = .075, and SRMR

= .037). All factor loadings were significant (*p*s < .001) and all were above .71. The four causal indicators accounted for approximately 73% of the variance in reading comprehension (Figure 2). Decoding and verbal reasoning accounted for unique reading comprehension variance after controlling for the other predictors in the model (2% and 17.3%, respectively). In separate models without the other predictors included, each of the causal indicators accounted for significant reading comprehension variance: decoding (β = .588, $p < .001$, $R^2 = .346$), verbal reasoning (β = .834, $p < .001$, $R^2 = .695$), nonverbal reasoning (β = .655, $p < .001$, $R^2 = .429$), and working memory (β = .370, $p < .001$, $R^2 = .137$). Similar to third grade, verbal reasoning emerged as the most important predictor of seventh grade reading comprehension; however, all constructs accounted for significant reading comprehension variance when included in separate models.

Tenth grade—The 4-factor MIMIC model provided nearly a perfect fit to the tenth grade sample ($\chi^2(34) = 29.07$, $p = .708$, CFI = 1.00 TLI = 1.01, RMSEA = .000, and SRMR = .023). All factor loadings were significant (*p*s < .001) and all were above .61. The four causal indicators accounted for approximately 87% of the variance in reading comprehension (Figure 3). Verbal reasoning was the only predictor that accounted for unique reading comprehension variance after controlling for the other predictors in the model (25.6%). Once again, in separate models without the other predictors included, each of the causal indicators contributed significant variance to reading comprehension: decoding (β = .492, $p < .001$, $R^2 = .242$), verbal reasoning (β = .921, $p < .001$, $R^2 = .849$), nonverbal reasoning (β = .738, $p < .001$, $R^2 = .544$), and working memory (β = .419, $p < .001$, $R^2 = .175$). Overall, verbal reasoning was the most important predictor of tenth grade reading comprehension; however, all constructs accounted for significant reading comprehension variance when included in separate models.

Multiple Group MIMIC Models of Reading Comprehension

Our third research question addressed fitting a multiple group model and assessing if the contributions of the four causal indicators to reading comprehension vary by grade level. In order to evaluate fit adequacy of a multiple group MIMIC model, measurement invariance (MI) of the structural components needs to be established first. Testing different levels of MI proceeds in a stepwise fashion, with subsequent levels indicating stronger MI (more constrained structural components). For the 4-factor MIMIC models in the current study, a fully invariant model would constrain all factor structures (factor loadings, factor variances and covariances, correlations, and residual variances) across the three grades to be equal. We established measurement invariance at the level of constrained factor loadings, factor variances and covariances, and correlations (our baseline model). Thus, the only cross-group constraint we could not impose was on the residual variances (see Appendix A for a complete step-by-step description of how MI was established and a table of model fit statistics at each invariance level). This level of MI allowed us to assume measurement stability (i.e., that we are assessing the same constructs across grade levels) and structural stability (i.e., the correlations among the four constructs are similar across grade levels).

Once we established a baseline model of the constrained structural components, in accordance with Treiblmaier et al. (2011) we conducted a canonical correlation analysis

(CCA) utilizing SAS 9.2 (SAS Institute Inc., 2012) as an empirical check that our four causal indicators were appropriately identified and accounting for the reading comprehension variance. To do this, we created two sets of maximally correlated linear composites from the observed indicators of our four predictors. Set A included the TOWRE SWE, a listening comprehension composite score, WASI Similarities, WASI Block Design, and reading span tasks. Set B included the TOWRE PDE, WASI Vocabulary, WASI Matrix Reasoning, and listening span tasks. As expected, we obtained a high correlation of .88 between the two sets of composites (averaged across our three grade levels). Next, we correlated the composites with our reading comprehension factor scores. Across the grade levels, we obtained an average correlation of .79 between the composites and reading comprehension factor scores (ranging from .73 to .83). The high correlation between the composites and reading comprehension factor scores indicate that our causal indicators are appropriately specified and account for the variance in reading comprehension at each grade level (Treiblmaier et al., 2011).

Based on our strong correlation from our empirical check, we were justified in testing the stability of the contributions of the four causal indicators to predicting reading comprehension across the three grade levels. To do this, we constrained each causal indicator to be equal across grades to test for model invariance. Chi-square difference tests indicated non-significant differences between the baseline model and models that constrained verbal reasoning, nonverbal reasoning, and working memory causal indicators to be equal across the three grades ($\chi^2(2) = 2.47$, $p = .291$; $\chi^2(2) = 0.141$, $p = .932$; $\chi^2(2) = 2.85$, $p = .241$, respectively) (Table 5). Decoding was the only causal indicator that could not be constrained to be equal across all grades ($\chi^2(2) = 11.31$ $p = .003$) (Table 5). However, constraining seventh and tenth grade decoding to be equal and freeing up third grade decoding resulted in a non-significant difference when compared to the baseline model ($\chi^2(1) = 2.29$ $p = .130$) (Table 5). Finally, we computed a model that constrained all causal indicators to be equal across grades (with the exception of third grade decoding) and found no significant chi-square difference between this model and the baseline model ($\chi^2(7) = 12.31$, $p = .091$) (Table 5). In sum, there is considerable stability in the contributions of verbal reasoning, nonverbal reasoning, and working memory to reading comprehension across the three grade levels. The contribution of decoding to reading comprehension is higher in third grade and then remains relatively stable between seventh and tenth grade.

Discussion

The current study illustrates the utility of using MIMIC models to investigate important predictors of reading comprehension between and across three distinct developmental grade levels. Our first two research questions addressed the adequacy of fitting 4-factor MIMIC models separately by grade and assessed the joint and unique variance estimates of the predictors to reading comprehension. At each grade level, the 4-factor MIMIC model with causal indicators of decoding, verbal reasoning, nonverbal reasoning, and working memory provided adequate fit. The causal indicators accounted for a substantial portion of the reading comprehension variance (73% to 87%) and verbal reasoning was found to be the most important predictor at each grade level (accounting for between 9.5% and 25.6% unique variance). Decoding also contributed unique variance to both third and seventh grade reading comprehension (5.7% and 2%, respectively). Our final research question addressed the stability of the magnitudes and contributions of the four causal indicators to reading comprehension across the three grade levels. A multiple group 4-factor MIMIC model with constrained structural components (baseline model) was compared to models that constrained each of the four causal indicators to be equal across grade levels. Verbal reasoning, nonverbal reasoning, and working memory exhibited relatively stable magnitudes and contributions to reading comprehension across development (grade levels). Decoding was the only causal indicator that could not be fully constrained across grades because decoding contributed significantly more variance to reading comprehension in third grade than in seventh and tenth grade. We conclude by relating our findings to past research, discussing the implications of our findings, and the limitations and directions for future research.

Predictors of Reading Comprehension

Simple view of reading (SVR)—Our findings are partially consistent with past research on the SVR component skills (Gough & Tunmer, 1986; Hoover & Gough, 1990) and the variability in the contributions of these skills across development (Adolf et al., 2006; Diakidoy et al., 2005; Kershaw & Schatschneider, 2012; Tilstra et al., 2009). In accordance with the SVR framework, we found that decoding and linguistic comprehension (vocabulary and listening comprehension were components of our verbal reasoning factor) were important predictors of reading comprehension (Catts et al., 2005; Tilstra et al., 2009). Moreover, consistent with past research, decoding exhibited a stronger relationship with reading comprehension in our early grade level (third grade) as compared to our later grade levels (seventh and tenth grade) (Tilstra et al., 2009; Verhoeven & van Leeuwe, 2008). This finding was confirmed by our individual 4-factor MIMIC models in which decoding contributed 5.7% unique comprehension variance in third grade, 2% unique comprehension variance in seventh grade, and no unique comprehension variance in tenth grade. The multiple group model also supported a developmental shift in the contribution of decoding to reading comprehension across the three grades. Decoding skills in third grade could not be constrained to be equal across grade levels, indicating a significantly greater contribution of decoding to third grade comprehension as compared to seventh and tenth grade reading comprehension.

In contrast to previous research on the SVR, our multiple group model did not support a developmental shift in the contribution of linguistic comprehension (our verbal reasoning factor) to reading comprehension across grade levels (Diakidoy et al., 2005; Tilstra et al., 2009). Although the unique contribution of verbal reasoning to reading comprehension increased as a function of grade level (9.5%, 17.3%, 25.6%, respectively), the magnitudes of these contributions were not significantly different across grade levels. In other words, we were able to fully constrain verbal reasoning in our multiple group model, which indicates the relative stability of our verbal reasoning factor across development. Thus, the observed grade level fluctuations in the unique variance estimates of verbal reasoning may be merely an indicator of the decreased contribution of word-level decoding skills to reading comprehension (as opposed to an increased contribution of higher-order verbal reasoning skills). Because the current study utilized a verbal reasoning factor (instead of a linguistic comprehension factor) future research needs to evaluate the generalizability and robustness of this finding across multiple measures and multiple grade levels.

Verbal and nonverbal reasoning—Consistent with past research, the current study found that verbal and nonverbal reasoning were important predictors of reading comprehension (Adlof et al., 2010; Schatschneider, et al., 2007; Tighe & Schatschneider, 2014). Verbal reasoning was the most important predictor across the three grade levels. Although nonverbal reasoning did not account for unique reading comprehension variance, nonverbal reasoning contributed significantly to reading comprehension at each grade level in a model without other included predictors. Inconsistent with previous research, our multiple group model did not support a change in the contributions of verbal or nonverbal reasoning skills to reading comprehension as a function of grade level (Cain & Oakhill, 1998).

Past research has noted a developmental shift in reasoning skills because the unique variance estimates increased across grade levels (Schatschneider et al., 2007; Tighe & Schatschneider, 2014). Similarly, the current study did observe an increase in the unique variance estimates of verbal reasoning; however, the multiple group model allowed us to investigate the stability of these contributions across development. Thus, we determined that the contributions of verbal and nonverbal reasoning to reading comprehension are relatively consistent across grade levels. This finding was surprising given that the number of FCAT passages (one of our reading comprehension measures) that require higher-order reasoning skills increases from third to tenth grade (30% to 70%) (Torgesen, Nettles, Howard, & Winterbottom, 2005). However, the limited body of research on the contributions of higher-order inferential and reasoning skills (particularly nonverbal reasoning) to reading comprehension makes it difficult to assess the importance of these skills across development. Oakhill and Cain (2012) conducted a longitudinal investigation of reading development across seven to 11-year-old children and reported that early discourse skills (inference and integration, comprehension monitoring, and knowledge of text structure) contributed unique variance to later reading comprehension after controlling for word reading, verbal IQ, receptive vocabulary knowledge, and the autoregressive comprehension effect. This suggests that as children transition from learning to read to reading for comprehension, higher-order discourse skills impact children's reading comprehension. In the current study, this is consistent with our finding of the decrease in importance of word-level decoding skills between the third grade level and the seventh and tenth grade levels; however, we did not include discourse skills in predicting reading comprehension. Thus, future research should explore these discourse skills in conjunction with reasoning skills to investigate how these skills develop and influence reading comprehension across development (Oakhill & Cain, 2012; Oakhill et al., 2003).

Working memory—Parallel to the mixed results regarding the importance of working memory to reading comprehension, the current study determined that working memory did not contribute unique reading comprehension variance at any grade level (Cain et al., 2004; Goff et al., 2005; Seigneuric & Ehrlich, 2005; Tighe & Schatschneider, 2014). Further, we did not find evidence for a developmental change in the contribution of working memory to reading comprehension across the three grade levels. Seigneuric and Ehrlich (2005) reported an increase in the importance of working memory (controlling for decoding and vocabulary) to reading comprehension in first through third graders. The current study did not evaluate working memory in the early elementary school grades; thus, there may exist a developmental shift in this relationship in earlier grade levels.

Cain et al. (2004) reported that working memory explained unique variance in the reading comprehension skills of eight to 11 year olds independent of verbal IQ, vocabulary, and word reading skills. In contrast to the current study, this suggests that working memory exerts a direct influence on reading comprehension independent of verbal reasoning skills. However, Oakhill and Cain (2012) reported that working memory did not account for unique variance in this same group of students after controlling for word reading, vocabulary, verbal IQ, the autoregressive effect of comprehension, and discourse skills (inferencing and integration, comprehension monitoring, and knowledge of text structure). Instead, these researchers emphasized the importance of the discourse skills to explaining unique variance in later reading comprehension development. Working memory capacity underlies the ability to generate inferences, integrate background knowledge, and engage in comprehension monitoring in order to form a cohesive representation of text during reading (Cain et al., 2004; Oakhill & Cain, 2012). Thus, it is possible that working memory does not account for additional reading comprehension variance independent of these discourse skills. More longitudinal research is needed to evaluate the developmental trajectory of working memory and discourse skills and their relationship to reading comprehension across development.

Implications of the Findings

The findings have implications for researchers and educators. For researchers, the current study highlights the feasibility of applying a causal indicator modeling approach to identifying component skills of reading comprehension. Causal indicator modeling presents an alternative to the traditional CFA approach by reversing causal flow and allowing researchers to estimate the underlying causes of a latent variable (i.e., reading comprehension). Consistent with past research, our 4-factor MIMIC models accounted for a substantial portion of the reading comprehension variance (73% to 87%) across the three grade levels (Tighe, 2012; Wagner, 2013). Therefore, MIMIC models appear to be particularly robust across multiple predictor constructs, grade levels, and measures of reading comprehension. Further, we extended these models by estimating the stability of the constructs and the magnitudes of the predictor contributions to reading comprehension across grades. By establishing measurement invariance of the structural components, we were able to illustrate the stability of our observed measures across grade levels. Assessing the stability of the contributions of the four predictors to reading comprehension allowed us to identify important predictors across development. Our findings suggest that decoding and verbal reasoning may be important constructs to target in reading interventions with elementary school children; whereas verbal reasoning may be a more important construct to target in middle school and high school intervention research.

For educators, the current study has begun to shed light on core reading component skills across development that may have important implications for instructional practices. Much of the past research has investigated predictors of reading comprehension at a single grade level or longitudinally in a cluster of similar grade levels (e.g., early elementary school years) (Kim, Petscher, Schatschneider, & Foorman, 2010; Seigneuric & Ehrlich, 2005). The current study utilized a cross-sectional design; however, selected three distinct points across children's reading development (elementary, middle, and high school grade levels). Thus, our establishment of measurement invariance and relative stability of the predictor contributions to reading comprehension yields important implications. First, the tasks used to measure decoding, verbal reasoning, nonverbal reasoning, and working memory are assessing the same constructs at each grade level. This suggests that we are able to measure the same construct consistently across grade levels. Second, the four predictor constructs jointly accounted for a substantial amount of the reading comprehension variance (73% to 87%) at each grade level. This suggests that regardless of reading ability and the inclusion of additional component skills (i.e., metalinguistic skills, discourse skills, naming speed, fluency), these four skills remain core contributors to reading comprehension across development. Finally, the relative stability of the contributions of the four predictors to reading comprehension suggests that it is important for instructional practices to target all of these skills across grade levels. However, it may be particularly beneficial to target word-level, decoding skills in earlier elementary school years as compared to the middle and high school years.

Limitations and Future Directions

There are four limitations that are worth noting. First, the four causal indicators accounted for a large portion of the reading comprehension variance; however, an additional 13% to 27% remains unexplained across the three grade levels. It would be useful to explore additional predictors of reading comprehension to account for this leftover variance. For example, past research has identified comprehension monitoring (Cain et al., 2004; Oakhill & Cain, 2012), inferencing and integration (Oakhill & Cain, 2012), phonological skills (de Jong & van der Leij, 2002), fluency (Tighe & Schatschneider, 2014), and morphological awareness (Deacon & Kirby, 2004) as important predictors of reading comprehension across several grade levels. Additionally, past research has included separate vocabulary knowledge (Nagy, Abbott, Vaughn, & Vermeulen, 2003), verbal IQ (Cain et al., 2004; Oakhill & Cain, 2012), and listening comprehension (Vellutino et al., 2007) constructs in predicting reading comprehension. The current study loaded a composite listening comprehension score and two WASI subtests (vocabulary and similarities) onto a verbal reasoning latent factor. It may be beneficial to include separate listening comprehension, verbal IQ, and vocabulary knowledge factors. Future research should investigate these predictors to determine an optimal model of reading comprehension.

Second, our sample of tenth graders exhibited slightly below average performance (relative to normative standards) on some of the norm-referenced assessments (e.g., TOWRE SWE and PDE). Moreover, the seventh and tenth graders obtained similar raw scores on some of the assessments (e.g., reading and listening span tasks). Thus, it is possible that our observed stability estimates from our multiple group model could partially reflect the similarities in skill levels between our seventh and tenth graders. However, it is difficult to compare raw score performance between grade levels because different passages and items were given for some of the assessments (e.g., listening comprehension tasks). Despite this limitation, we were able to model stability (with the exception of decoding) across multiple measures and grade levels (third, seventh, and tenth). Future research should investigate the stability of the constructs in additional samples.

Third, our working memory construct only included two verbal working memory tasks. Both of these tasks were similar in format and relied exclusively on processing sentences and word recall. These measures may not have been exclusively tapping working memory abilities. Our listening span measure may have conflated working memory with listening comprehension skills and our reading span measure may have conflated working memory with decoding skills. Past research on working memory has identified that verbal and numerical working memory tend to be more highly correlated with reading comprehension skills than visuo-spatial working memory (Daneman & Merikle, 1996; Seigneuric et al., 2000). Future research may want to incorporate multiple types of verbal working memory assessments and/or include numerical working memory tasks to eliminate the potential confound with decoding skills.

Finally, our study is cross-sectional and therefore, it is difficult to assess developmental shifts in the contributions of predictors to reading comprehension. We chose three grades that represented discrete time points across development (i.e., elementary, middle, and high school years). However, it would be better for future research to investigate the contributions of component skills to reading comprehension longitudinally.

Conclusion

The current study illustrated the utility of applying a causal indicator modeling framework to investigate important predictors of reading comprehension in third, seventh, and tenth grade students. The causal indicators of decoding, verbal reasoning, nonverbal reasoning, and working memory accounted for a large proportion of the reading comprehension variance (73% to 87%) at each grade level. Verbal reasoning was identified as the most important predictor of reading comprehension across all grades. A multiple group 4-factor MIMIC model indicated relative stability in the contributions of verbal reasoning, nonverbal reasoning, and working memory to reading comprehension across the three grade levels. Decoding skills exhibited an increased contribution to third grade reading comprehension as compared to seventh and tenth reading comprehension skills. Future research should investigate additional component skills and examine these skills longitudinally to build the most comprehensive model of reading comprehension across children's development of reading skills.

Acknowledgments

Support for this manuscript was provided by grant number P50 HD052120 from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development awarded to the second and third authors. Support was also provided by Predoctoral Interdisciplinary Research Training Fellowships R305B04074 and R305B090021 from the Institute of Education Sciences awarded to the first author.

Appendix A: Establishing Measurement Invariance in Multiple Group Causal Indicator Models

Measurement invariance (MI) refers to scores on a latent construct having the same operational definition across different groups, time points, or methods of measurement administration (Meade & Lautenschlager, 2004). To establish MI across the structural components of our three independent groups (third, seventh, and tenth grade), we followed a six-step process and inspected and compared model fit indices at each step (Cheung & Rensvold, 2002; Kline, 2011). It is important to note that the steps proceed in hierarchical fashion such that subsequent levels impose additional equivalence constraints and therefore, indicate stronger MI.

Step one is to ensure that the same general model fits reasonably well at all grade levels. As demonstrated in our results section, a 4-factor MIMIC model provided adequate fit at each grade level. Inspecting the model fit indices and factor loadings of the observed measures onto the latent constructs also showed similar loadings across the three grades. Thus, we can explore building a multiple group MIMIC model.

Step two is to obtain a configural or baseline model (labeled M_0 in Table 6) by fitting the 4-factor MIMIC model across the three grades without any invariances (or cross-group equality constraints). By default for a multiple group model, Mplus imposes cross-group equality constraints on factor loadings and intercepts. We are unable to relax the cross-group equality constraints on intercepts because this would lead to an unidentified model. However, we were able to run a baseline model that freed up the default cross-group constraints on factor loadings. This model is referred to as a baseline model (M_0) because this model was compared with subsequent models in our later steps of testing for MI. Comparing chi-square values from the three models (one per grade) and M_0 demonstrated that the baseline model is merely the sum of chi-square values and degrees of freedom from each of the individual models.

Step three is to test for construct-level metric invariance or equal factor loadings. To achieve this, we constrained all unstandardized factor loadings to be equal across the three groups (Table 6). A chi-square difference test between this model (M_1) and the baseline model (M_0) revealed a non-significant difference ($M_1{-}M_0$ $\chi^2(12) = 15.39$, $p = .221$). Inspection of the model fit indices revealed an increase in RMSEA and TLI values in the constrained M_1 model, indicating better model fit. Therefore, we established invariance of factor loadings, which indicates that the latent constructs (decoding, verbal reasoning, nonverbal reasoning, and working memory) are equivalent across grades.

Step four is to test for the equivalence of construct variances and covariances. To do this, we imposed constraints on factor loadings and factor variances and covariances (Table 6). A chi-square difference test between this model (M_2) and the Step three model (M_1) revealed no significant differences ($M_2{-}M_1$ $\chi^2(8) = 1.37$, $p = .994$). The model fit indices show increased model fit for the RMSEA, TLI, and CFI values. Thus, cross-group equivalence of factor variances and covariances is supported.

Step five is to test for invariance of the correlations among our latent constructs (decoding, verbal reasoning, nonverbal reasoning, and working memory). To do this, cross-group equality constraints were imposed on factor loadings, factor covariances, and correlations (Table 6). A chi-square difference test between this model (M_3) and the model in Step four (M_2) revealed no significant differences ($M_3 - M_2\ \chi^2(12) = 18.05$, $p = .114$). Inspection of the model fit indices revealed increased model fit for the RMSEA and TLI values. Thus, cross-group equivalence of predictor correlations is supported.

Step six is to test for invariance of the residual variances. To do this, cross-group equality constraints were enforced on factor loadings, factor variances and covariances, construct correlations, and residual variances (Table 6). A chi-square difference test between this model (M_4) and the model in Step five (M_3) revealed a significant difference ($M_4 - M_3(22) = 48.76$, $p < .001$). Inspection of the model fit indices demonstrated decreased model fit for the RMSEA, CFI, and TLI values. Although model fit was still in an overall acceptable range (Hu & Bentler, 1998), we could not declare cross-group equivalence of residual variances. Thus, we utilized model M_3 as our baseline model to test for the stability of the estimates of the four causal indicator pathways.

References

Adlof SM, Catts HW, Lee J. Kindergarten predictors of second versus eighth grade reading comprehension impairments. Journal of Learning Disabilities. 2010; 43(4):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.

Alloway TP, Alloway RG. Investigating the predictive roles of working memory and IQ in academic attainment. Journal of Experimental Child Psychology. 2010; 106:20–29. [PubMed: 20018296]

Asbell S, Donders J, Van Tubbergen M, Warschausky S. Predictors of reading comprehension in children with cerebral palsy and typically developing children. Child Neuropsychology. 2010; 16:313–325. [PubMed: 20455127]

Baddeley, AD.; Hitch, G. Working memory. In: Bower, GH., editor. Psychology of learning and motivation. Vol. 8. New York, NY: Academic Press, Inc; 1974. p. 47-87.

Beck, I.; McKeown, M. Conditions of vocabulary acquisition. In: Barr, R.; Kamil, M.; Mosenthal, P.; Pearson, PD., editors. Handbook of reading research. Vol. 2. Mahwah, NJ: Lawrence Erlbarum Associates; 1991. p. 789-814.

Bollen KA, Bauldry S. Three Cs in measurement models: Causal indicators, composite indicators, and covariates. Psychological Methods. 2011; 16(3):265–284. [PubMed: 21767021]

Bollen KA, Davis WF. Causal indicator models: Identification, estimation, and testing. Structural Equation Modeling: A Multidisciplinary Journal. 2009; 16(3):498–522.

Cain, K.; Oakhill, J. Comprehension skill and inference-making ability: Issues of causality. In: Hulme, C.; Joshi, R., editors. Reading and spelling: Development and disorders. Mahwah, NJ: Lawrence Erlbaum Associates; 1998. p. 329-342.

Cain, K.; Oakhill, J., editors. Children's comprehension problems in oral and written language: A cognitive perspective. New York, NY: The Guilford Press; 2007.

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(1):31–42.

Catts, HW.; Hogan, TP.; Adlof, SM. Developmental changes in reading and reading disabilities. In: Catts, HW.; Kamhi, AG., editors. Connections between language and reading disabilities. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2005. p. 2025-2040.

Chen R, Vellutino FR. Prediction of reading ability: A cross-validation study of the simple view of reading. Journal of Literacy Research. 1997; 29(1):1–24.

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

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(3):277–299.

Daneman M, Carpenter PA. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior. 1980; 19:450–466.

Daneman M, Merikle PM. Working memory and language comprehension: A metaanalysis. Psychonomic Bulletin & Review. 1996; 3(4):422–433. [PubMed: 24213976]

Deacon SH, Kirby JR. Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. Applied Psycholinguistics. 2004; 25:223–238.

de Jong PF, van der Leij A. Effects of phonological abilities and linguistic comprehension on the development of reading. Scientific Studies of Reading. 2002; 6(1):51–77.

Diakidoy IN, Stylianou P, Karefillidou C, Papageorgiou P. The relationship between listening and reading comprehension of different types of text at increasing grade levels. Reading Psychology. 2005; 26:55–80.

Dreyer LG, Katz L. An examination of "the simple view of reading. National Reading Conference Yearbook. 1992; 41:169–175.

Florida Department of Education. Florida Comprehensive Assessment Test: Technical Report 2006. Tallahassee, FL: 2006. Retrieved March 15, 2012, from http://fcat.fldoe.org/pdf/fc06tech.pdf

Fuchs D, Compton DL, Fuchs LS, Bryant VJ, Hamlett CL, Lambert W. First-grade cognitive abilities as long-term predictors of reading comprehension and disability status. Journal of Learning Disabilities. 2012; 45(3):217–231. [PubMed: 22539057]

Gaulin CA, Campbell TF. Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. Perceptual and Motor Skills. 1994; 79(1):55–64. [PubMed: 7991333]

Goff D, Pratt C, Ong B. The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. Reading and Writing: An Interdisciplinary Journal. 2005; 2:127–160.

Gough PB, Tunmer WE. Decoding, reading, and reading disability. RASE: Remedial and Special Education. 1986; 7(1):6–10.

Harcourt Brace. Stanford Achievement Test. Ninth. San Antonio, TX: Author; 1997.

Hoover WA, Gough PB. The simple view of reading. Reading and Writing: An Interdisciplinary Journal. 1990; 2:127–160.

Hu LT, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods. 1998; 3(4):424–453.

Jensen, AR. Bias in mental testing. New York, NY: Free Press; 1980.

Kershaw S, Schatschneider C. A latent variable approach to the simple view of reading. Reading and Writing: An Interdisciplinary Journal. 2012; 25(2):433–464.

Kim YS, Petscher Y, Schatschneider C, Foorman B. Does growth rate in oral reading fluency matter in predicting reading achievement? Journal of Educational Psychology. 2010; 102(3):652–667.

Kline, RB. Principles and practice of structural equation modeling: Third edition. New York, NY: The Guilford Press; 2011.

McVay J, Kane MJ. Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. Journal of Experimental Psychology: General. 2011; 141(2):302–320. [PubMed: 21875246]

Meade AW, Lautenschlager GJ. A comparison of item response theory and confirmatory factor analytic tests of measurement invariance. Journal of Applied Psychology. 2004; 93:568–592. [PubMed: 18457487]

Molloy, PJ. (Doctoral Dissertation). Available from Dissertation Abstracts International, 2072A. (UMI No. AAMNN19626). 1997. The role of individual differences in working memory in reading and listening comprehension in intermediate grade students.

Muthén, LK.; Muthén, BO. Mplus User's Guide. Sixth. Los Angeles, CA: Muthén & Muthén; 1998–2010.

Nagy WE, Berninger V, Abbott R, Vaughan K, Vermeulen K. Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. Journal of Educational Psychology. 2003; 95:730–742.

Nation, K. Reading comprehension and vocabulary: What's the connection?. In: Wagner, RK.; Schatschneider, C.; Phythian-Sence, C., editors. Beyond decoding: The behavioral and biological foundations of readng comprehension. New York, NY: The Guilford Press; 2009. p. 176-194.

Nation K, Adams JW, Bowyer-Crane CA, Snowling MJ. Working memory deficits in poor comprehenders reflect underlying language impairments. Journal of Experimental Child Psychology. 1999; 73(2):139–158. [PubMed: 10328862]

Oakhill JV, Cain K. The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. Scientific Studies of Reading. 2012; 16(2):91–121.

Oakhill JV, Cain K, Bryant PE. The dissociation of single-word reading and text comprehension: Evidence from component skills. Language and Cognitive Processes. 2003; 18:443–468.

Psychological Corporation. Wechsler Abbreviated Scale of Intelligence (WASI) manual. San Antonio, TX: Author; 1999.

SAS Institute Inc. Base SAS 9.2 utilities: Reference. Cary, NC: Author; 2012.

Schatschneider, C.; Harrell, E.; Buck, J. An individual differences approach to the study of reading comprehension. In: Wagner, RK.; Muse, AE.; Tannenbaum, KR., editors. Vocabulary acquisition: Implications for reading comprehension. New York, NY: The Guilford Press; 2007. p. 249-275.

Seigneuric A, Ehrlich MF. Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. Reading and Writing: An Interdisciplinary Journal. 2005; 18(7–9):617–656.

Seigneuric A, Ehrlich MF, Oakhill J, Yuill N. Working memory resources and children's reading comprehension. Reading and Writing: An Interdisciplinary Journal. 2000; 18:617–656.

Sternberg RJ, Powell JS. Comprehending verbal comprehension. American Psychologist. 1983; 38:878–893.

Sticht, TG.; James, JH. Listening and reading. In: Pearson, PD.; Barr, R.; Kamil, ML.; Mosenthal, P., editors. Handbook of reading research. New York, NY: Longman; 1984. p. 255-292.

Swanson HL, Berninger V. The role of working memory in skilled and less skilled readers' comprehension. Intelligence. 1995; 21(1):83–108.

Swanson HL, Howell M. Working memory, short-term memory, and speech rate as predictors of children's reading performance at different ages. Journal or Educational Psychology. 2001; 93:720–734. [PubMed: 11477585]

Swanson HL, O'Connor R. The role of working memory and fluency practice on the reading comprehension of students who are dysfluent readers. Journal of Learning Disabilities. 2009; 42(6):548–575. [PubMed: 19745196]

Tighe, EL. Unpublished masters thesis. Florida State University; 2012. An investigation of the dimensionality of morphological and vocabulary knowledge in Adult Basic Education students.

Tighe EL, Schatschneider C. A dominance analysis approach to determining predictor importance in third, seventh, and tenth grade reading comprehension skills. Reading and Writing: An Interdisciplinary Journal. 2014; 27(1):101–127.

Tilstra J, McMaster K, Van den Broek P, Kendeou P, Rapp D. Simple but complex: Components of the simple view of reading across grade levels. Journal of Research in Reading. 2009; 32(4):383–401.

Tiu RD Jr, Thompson LA, Lewis BA. The role of IQ in a component model of reading. Journal of Learning Disabilities. 2003; 36(5):424–436. [PubMed: 15497486]

Torgesen, J.; Nettles, S.; Howard, P.; Winterbottom, R. Brief report of a study to investigate the relationship between several brief measure of reading fluency and performance on the Florida Comprehensive Assessment Test-Reading in 4th, 6th, 8th, and 10th grades. Tallahassee, FL:

Florida Center for Reading Research; 2005. Technical Report #6Retrieved from: http://www.fcrr.org/TechnicalReports/Progress_monitoring_report.pdf

Torgesen, JK.; Wagner, RK.; Rashotte, CA. Test of Word Reading Efficiency. Austin, TX: Pro-Ed; 1999.

Treiblmaier H, Bentler PM, Mair P. Formative constructs implemented via common factors. Structural Equation Modeling. 2011; 18:1–17.

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(1): 3–32.

Verhoeven L, van Leeuwe J. Prediction of the development of reading comprehension: A longitudinal study. Applied Cognitive Psychology. 2008; 22:407–423.

Wagner, RK. A causal indicator model of reading comprehension; Paper presented at the Frontier Lecture Series at Texas A&M University; College Station, Texas. 2013 Jan.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript

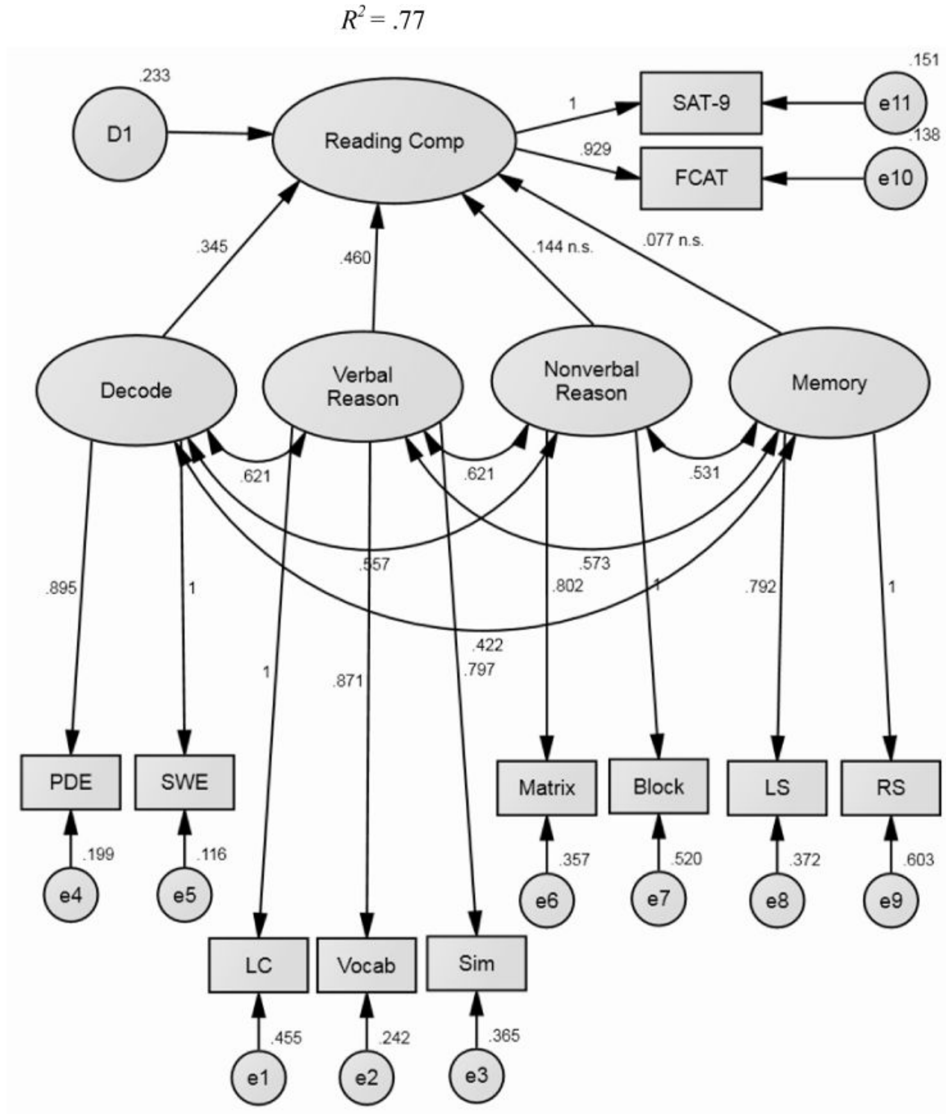

Figure 1.
4-Factor MIMIC Model of Third Grade Reading Comprehension
Note: This model presents standardized parameter estimates. All loadings are significant at $p < .001$ unless otherwise specified (n.s = not significant.).

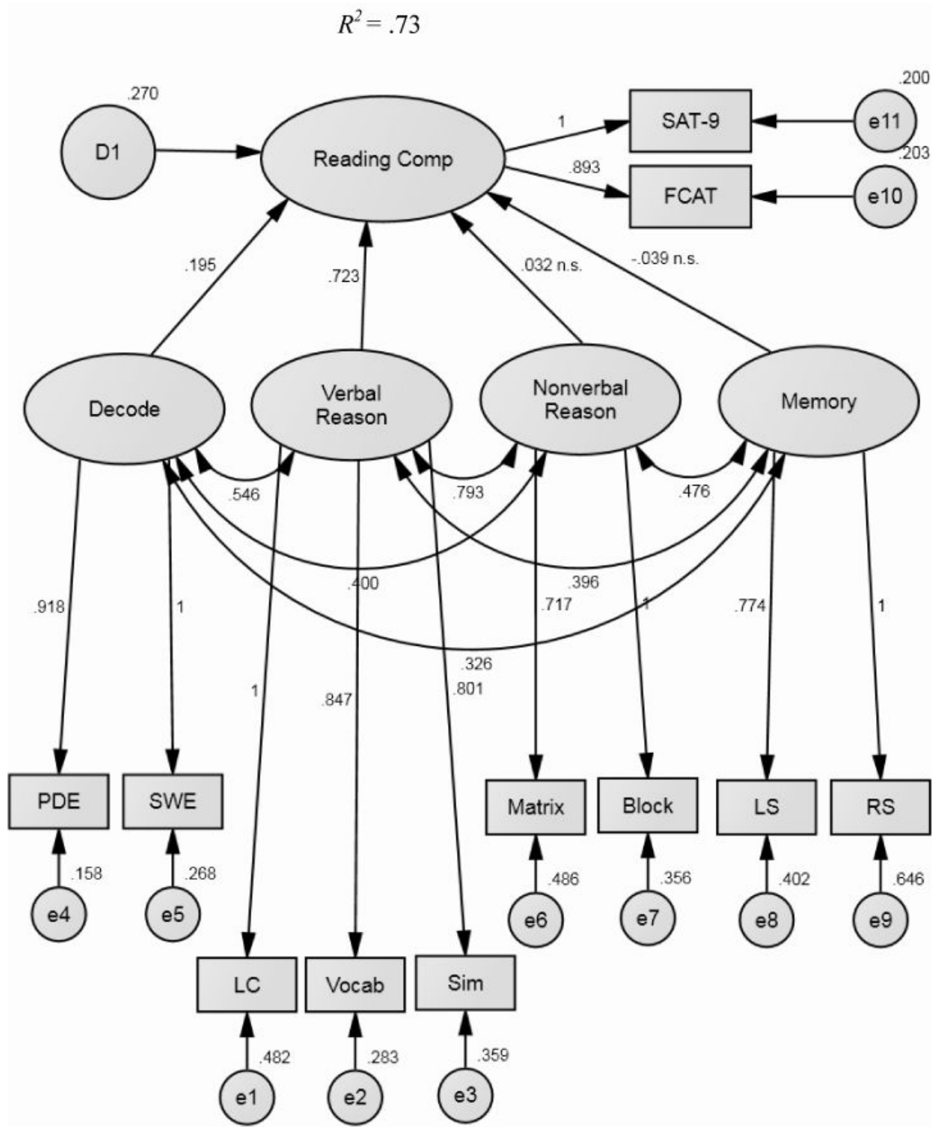

Figure 2.

4-Factor MIMIC Model of Seventh Grade Reading Comprehension

Note: This model presents standardized parameter estimates. All loadings are significant at p< .001 unless otherwise specified (n.s. = not significant).

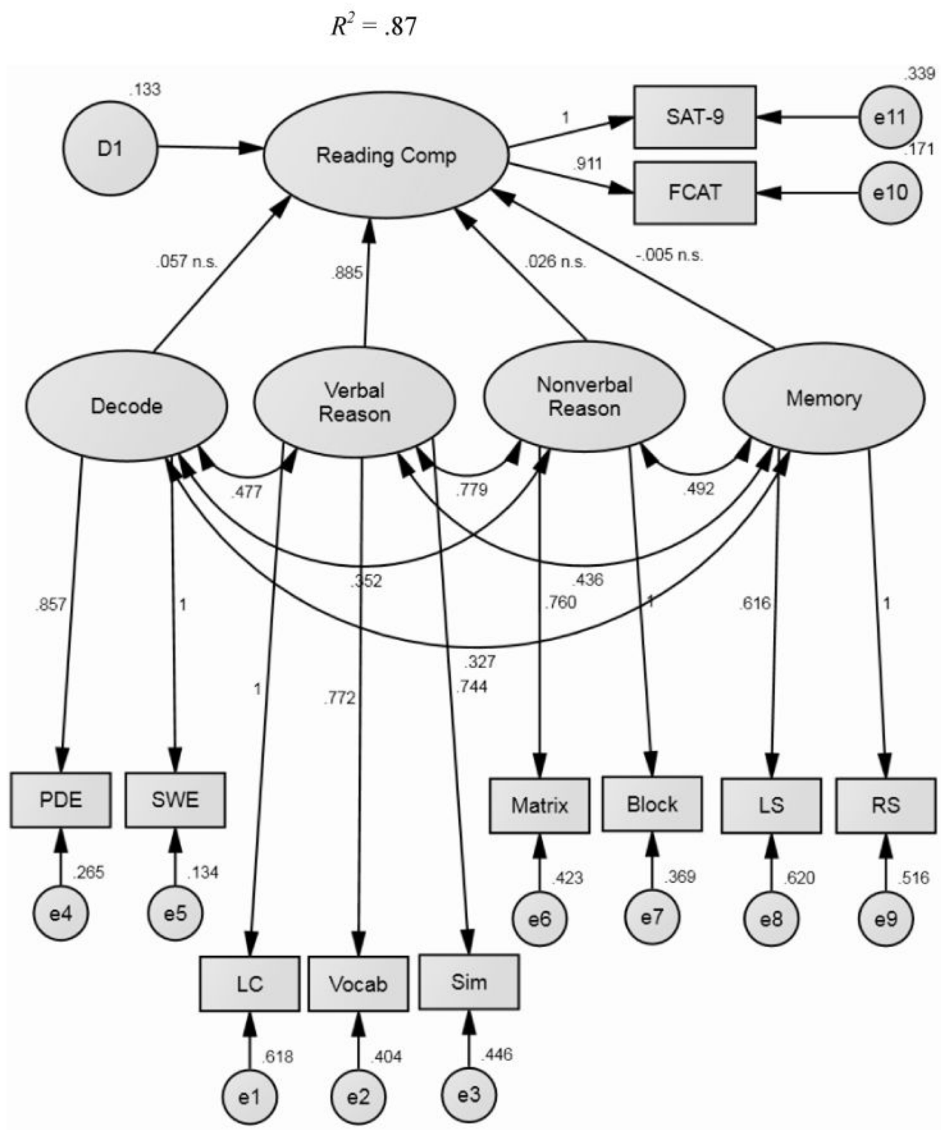

Figure 3.

4-Factor MIMIC Model of Tenth Grade Reading Comprehension

Note: This model presents standardized parameter estimates. All loadings are significant at $p < .001$ unless otherwise specified (n.s. = not significant).

Table 1

Means and Standard Deviations for all Measures with Third, Seventh, and Tenth Graders

	Third Grade		**Seventh Grade**		**Tenth Grade**	
Measure	***N***	***M(SD)***	***N***	***M(SD)***	***N***	***M(SD)***
Reading Comprehension						
FCAT	207	310.59(62.63)	184	319.89(50.80)	180	307.63(43.28)
SAT-9	210	638.00(47.13)	184	696.38(34.20)	180	696.33(32.02)
Decoding						
TOWRE SWE	215	103.34(14.56)	188	102.05(11.87)	182	91.96(10.33)
TOWRE PDE	215	100.49(16.03)	188	100.60(14.81)	182	91.17(16.19)
Working Memory						
Reading Span	214	19.22(8.05)	186	25.80(6.54)	182	24.87(7.16)
Listening Span	215	19.68(7.89)	187	25.18(7.19)	182	24.94(7.85)
Verbal Reasoning						
Listening Passage 1	214	4.86(1.60)	187	3.71(1.21)	182	2.20(1.00)
Listening Passage 2	215	2.48(1.28)	187	3.37(1.46)	182	2.05(0.82)
Listening Passage 3	214	2.84(1.14)	186	4.23(1.45)	182	3.53(1.17)
LC Composite	215	3.39 (1.05)	187	3.77(1.07)	182	2.60 (0.73)
WASI Vocabulary	210	48.71(11.81)	184	48.53(9.25)	177	46.40(9.27)
WASI Similarities	210	53.28(11.15)	186	49.39(11.34)	182	47.34(9.60)
Nonverbal Reasoning						
WASI Matrix Reasoning	210	52.05(10.87)	186	49.93(8.98)	182	47.03(8.97)
WASI Block Design	213	49.81(10.62)	186	49.48(10.94)	182	47.43(10.35)

Note: FCAT = Florida Comprehensive Assessment Test. SAT-9 = Stanford Achievement Test-Ninth Edition. TOWRE = Test of Word Reading Efficiency. SWE = Sight Word Efficiency. PDE = Phonemic Decoding Efficiency. WASI = Wechsler Abbreviated Scale of Intelligence.

Table 2

Correlations Among All Third Grade Measures

Measure	1	2	3	4	5	6	7	8	9	10	11
1. SAT-9	–	.86	.64	.61	.41	.28	.64	.63	.56	.41	.50
2. FCAT	–	–	.65	.62	.43	.33	.65	.62	.55	.44	.49
3. SWE	–	–	–	.84	.34	.20	.41	.52	.47	.36	.42
4. PDE	–	–	–	–	.30	.23	.32	.52	.45	.39	.39
5. WM LS	–	–	–	–	–	.50	.40	.40	.28	.27	.35
6. WM RS	–	–	–	–	–	–	.28	.36	.27	.26	.25
7. LC	–	–	–	–	–	–	–	.61	.55	.34	.44
8. Vocab	–	–	–	–	–	–	–	–	.74	.38	.39
9. Sim	–	–	–	–	–	–	–	–	–	.32	.42
10. Block	–	–	–	–	–	–	–	–	–	–	.56
11. Matrix	–	–	–	–	–	–	–	–	–	–	–

Note: N = 215. SAT-9 = Stanford Achievement Test-Ninth Edition. FCAT = Florida Comprehensive Assessment Test. SWE = Sight Word Efficiency. PDE = Phonemic Decoding Efficiency. WM LS = Working Memory Listening Span. WM RS = Working Memory Reading Span. LC = Listening Comprehension. Vocab = Vocabulary Sim = Similarities. All correlations are significant at $p< .01$.

Table 3

Correlations Among All Seventh Grade Measures

Measure	1	2	3	4	5	6	7	8	9	10	11
1. SAT-9	–	.80	.49	.49	.22	.18	.62	.63	53	.48	.44
2. FCAT	–	–	.43	.48	.24	.17	.66	.63	.54	.47	.42
3. SWE	–	–	–	.79	.21	.23	.35	.39	.33	.24	.32
4. PDE	–	–	–	–	.20	.21	.33	.43	.44	.26	.30
5. WM LS	–	–	–	–	–	.46	.32	.27	.22	27	.32
6. WM RS	–	–	–	–	–	–	.26	.14[a]	.09[a]	.17	.26
7. LC	–	–	–	–	–	–	–	.57	.53	.49	.40
8. Vocab	–	–	–	–	–	–	–	–	.73	.50	.46
9. Sim	–	–	–	–	–	–	–	–	–	.57	.45
10. Block	–	–	–	–	–	–	–	–	–	–	.58
11. Matrix	–	–	–	–	–	–	–	–	–	–	–

Note: N = 188. SAT-9 = Stanford Achievement Test-Ninth Edition. FCAT = Florida Comprehensive Assessment Test. SWE = Sight Word Efficiency. PDE = Phonemic Decoding Efficiency. WM LS = Working Memory Listening Span. WM RS = Working Memory Reading Span. LC = Listening Comprehension. Vocab = Vocabulary. Sim = Similarities. Correlations are significant at $p < .05$.

[a]correlations are not significant $p > .05$.

Table 4

Correlations Among All Tenth Grade Measures

Measure	1	2	3	4	5	6	7	8	9	10	11
1. SAT-9	–	.74	.37	.35	.22	.29	.52	.55	.54	.50	.48
2. FCAT	–	–	.41	.38	.18	.27	.52	.65	.65	.51	.51
3. SWE	–	–	–	.80	.16	.22	.28	.41	.27	.26	.25
4. PDE	–	–	–	–	.19	.19	.26	.32	.25	.26	.20
5. WM LS	–	–	–	–	–	.43	.15	.25	.19	.24	.25
6. WM RS	–	–	–	–	–	–	.26	.22	.18	.27	.24
7. LC	–	–	–	–	–	–	–	.47	.42	.42	.36
8. Vocab	–	–	–	–	–	–	–	–	.60	.49	.45
9. Sim	–	–	–	–	–	–	–	–	–	.45	.43
10. Block	–	–	–	–	–	–	–	–	–	–	.60
11. Matrix	–	–	–	–	–	–	–	–	–	–	–

Note: N = 182. SAT-9 = Stanford Achievement Test-Ninth Edition. FCAT = Florida Comprehensive Assessment Test. SWE = Sight Word Efficiency. PDE = Phonemic Decoding Efficiency. WM LS = Working Memory Listening Span. WM RS = Working Memory Reading Span. LC = Listening Comprehension. Vocab = Vocabulary. Sim = Similarities. All correlations are significant at $p < .05$.

Table 5

Testing the Measurement Invariance of Predictors of Reading Comprehension Across Three Grades

Model	**χ^2(df)**	***P***	***P* χ^2 diff**	**CFI**	**TLI**	**RMSEA**	**SRMR**
Baseline	240.005(146)	<.001	—	.971	.968	.057	.051
All Equal Decode	251.518(148)	<.001	< .01	.968	.965	.060	.055
Decode (freeing up 3rd grade)	242.297(147)	<.001	>. 05	.971	.967	.058	.052
All Equal Verbal	242.479(148)	<.001	> .05	.970	.967	.058	.051
All Equal Nonverbal	240.146(148)	<.001	> .05	.972	.969	.057	.051
All Equal Working Memory	242.851(148)	<.001	> .05	.971	.968	.057	.052
All Equal Predictors except 3rd grade Decode	252.310(153)	<.001	> .05	.970	.967	.058	.053
All Equal Predictors	257.437(154)	<.001	< .05	.968	.966	.059	.056

* *Note*. Only significant changes occur when 3rd grade decoding is constrained.

Table 6

Testing for Multiple Group Measurement Invariance of the Structural Components

Model	χ^2(df)	*p*	χ^2 diff	CFI	TLI	RMSEA	SRMR
Baseline (No Constraints) M_0	205.194(114)	<.001	—	.972	.960	.064	.036
Equal Factor Loadings (M_1)	220.583(126)	<.001	> .05	.971	.962	.062	.040
Equal Variances/Covariances (M_2)	221.954(134)	<.001	>.05	.973	.967	.058	.042
Equal Correlations (M_3)	240.005(146)	<.001	>.05	.971	.968	.057	.051
Equal Residual Variances (M_4)	288.765(168)	<.001	<.01	.963	.964	.061	.057

Note: At each subsequent model, all of the previous model constraints are already implied.