2008

Relationships Between Measures of Word Knowledge and Reading Comprehension in Third- and Seventh-Grade Children

Kendra R. Tannenbaum
RELATIONSHIPS BETWEEN MEASURES OF WORD KNOWLEDGE AND
READING COMPREHENSION IN THIRD- AND SEVENTH-GRADE CHILDREN

By

KENDRA R. TANNENBAUM

A Dissertation submitted to the
Department of Psychology
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Summer Semester, 2008
The members of the Committee approve the dissertation of Kendra R. Tannenbaum defended on June 26, 2008.

___________________________________
Joseph K. Torgesen
Professor Directing Dissertation

___________________________________
Stephanie Al Otaiba
Outside Committee Member

___________________________________
Janet Kistner
Committee Member

___________________________________
Christopher J. Lonigan
Committee Member

___________________________________
Richard K. Wagner
Committee Member

The Office of Graduate Studies has verified and approved the above named committee members.
ACKNOWLEDGEMENTS

I would like to acknowledge several people who provided support and guidance during my time as a graduate student at Florida State University. First and foremost, I thank my major professor, Joseph Torgesen. This project would not have been possible without his financial and emotional support. Not only has he served as an advisor, he has taught me important lessons about professionalism, leadership, and ethical research practices. I also thank the members of my doctoral committee, Drs. Richard Wagner, Chris Lonigan, Janet Kistner, and Stephanie Al Otaiba, for their helpful suggestions. I would also like to express my gratitude to the undergraduate students who assisted in the completion of this project.

My graduate school experience has been enriched by the wonderful lifelong friendships that have been developed. I thank all of my classmates and other students in the Clinical Psychology program. Without them, this process would not have been as enjoyable. In particular, I thank Lara Jakobsons for her unwavering support as both a friend and labmate. I also thank the many friends from high school and college who have been so caring and encouraging over the years.

It is important that I also recognize the individuals at Children’s Hospitals and Clinics of Minnesota who have guided me through the final stage of the doctoral process. I will always be grateful to my supervisors and fellow interns for their words of encouragement and compassionate mentoring. Many thanks to Dr. Sharon Berry for helping me find the right path in life.

Finally, I want to extend a very special thank you to my family. Everything that I have accomplished has been possible because of my parents. They have celebrated with me during times of joy and provided words of wisdom to guide me through difficult times. Most importantly, they have given me endless love and support.
# TABLE OF CONTENTS

List of Tables ........................................................................................................ v
List of Figures ....................................................................................................... vi
Abstract................................................................................................................ viii

INTRODUCTION................................................................................................. 1

METHOD ............................................................................................................. 10

RESULTS............................................................................................................. 22

DISCUSSION....................................................................................................... 54

APPENDICES ...................................................................................................... 62

REFERENCES...................................................................................................... 70

BIOGRAPHICAL SKETCH ................................................................................. 77
LIST OF TABLES

1. Comparison of Mean Scores on the FCAT-SSS between Third- and Seventh-Grade Students in the Current Samples and in the School District..... 20

2. Comparison of School-Level Mean Scores on the FCAT-SSS between Participants in the Current Study and the Entire Grade ................................... 21

3. Descriptive Statistics and Intercorrelations for All Observed Variables, 3rd Grade ........................................................................................................ 30

4. Descriptive Statistics and Intercorrelations for All Observed Variables, 7th Grade ........................................................................................................ 31

5. Model Fit Indices, 3rd Grade ........................................................................... 32

6. Model Fit Indices, 7th Grade ........................................................................... 33
# LIST OF FIGURES

1. Three Factor Measurement Model of Breadth, Depth, and Fluency in the Third-Grade Sample ....................................................................................... 34

2. Two Factor Measurement Model of Fluency and Breadth/Depth in the Third-Grade Sample ....................................................................................... 35

3. Two Factor Measurement Model of Breadth and Depth/Fluency in the Third-Grade Sample ....................................................................................... 36

4. Two Factor Measurement Model of Depth and Breadth/Fluency in the Third-Grade Sample ....................................................................................... 37

5. One Factor Measurement Model of General Word Knowledge in the Third-Grade Sample ....................................................................................... 38

6. Three Factor Measurement Model of Breadth, Depth, and Fluency in the Seventh-Grade Sample ................................................................................... 39

7. Two Factor Measurement Model of Fluency and Breadth/Depth in the Seventh-Grade Sample ................................................................................... 40

8. Two Factor Measurement Model of Breadth and Depth/Fluency in the Seventh-Grade Sample ................................................................................... 41

9. Two Factor Measurement Model of Depth and Breadth/Fluency in the Seventh-Grade Sample ................................................................................... 42

10. One Factor Measurement Model of General Word Knowledge in the Seventh-Grade Sample .................................................................................... 43

11. One Factor Structural Model in the Third-Grade Sample ......................... 44

12. One Factor Structural Model with Processing Speed in the Third-Grade Sample ........................................................................................................ 45

13. One Factor Structural Model with FCAT-SSS in the Third-Grade Sample ...... 46

14. One Factor Structural Model with SAT-10 in the Third-Grade Sample ........... 47

15. Two Factor Structural Model in the Seventh-Grade Sample ....................... 48

16. Two Factor Structural Model with Processing Speed in the Seventh-Grade Sample ........................................................................................................ 49
17. Structural Model with Vocabulary Fluency as the Only Predictor Variable in the Seventh-Grade Sample ................................................................. 50

18. Structural Model with Vocabulary Fluency and Processing Speed as Predictor Variables in the Seventh-Grade Sample ......................................................... 51

19. Two Factor Structural Model with FCAT-SSS in the Seventh-Grade Sample .......................................................................................................................... 52

20. Two Factor Structural Model with SAT-10 in the Seventh-Grade Sample...... 53
This study examined the relationships between reading comprehension and breadth, depth, and fluency of word knowledge in third- and seventh-grade students. The Expressive One-Word Picture Vocabulary Test-Third Edition (EOWPVT) and the Receptive One-Word Picture Vocabulary Test-Second Edition (ROWPVT) were used to measure breadth of word knowledge. The Multiple Contexts subtest of the Test of Word Knowledge (TOWK) and the Associations subtest of The WORD Test-2 were used to measure depth of word knowledge. The Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities and the Picture Naming: Nouns subtest of the Test of Word Finding-2 (TWF-2) and Test of Adolescent/Adult Word Finding (TAWF) were used to measure fluency of word knowledge.

Confirmatory factor analyses showed that a one factor model of General Word Knowledge provided the best fit to the data in the third-grade sample. A two-factor model of Fluency and Breath/Depth emerged as the best fitting model in the seventh-grade sample, and Breadth/Depth had a stronger relationship to Reading Comprehension than did Fluency. Equivalence testing indicated that the two measures of reading comprehension were invariant across the two groups; however, the majority of the measures of word knowledge were not invariant. The results of the current study, in combination with the results of an earlier study conducted by Tannenbaum, Torgesen, & Wagner (2006), do not support a robust three-factor model of word knowledge.
INTRODUCTION

One of the most consistent findings in reading research is the strong relationship between word knowledge and reading comprehension (see Anderson & Freebody, 1981; Baumann, Kame’enui, & Ash, 2003; Graves, 1986; Stahl & Fairbanks, 1986, for review). The National Reading Panel (2000) concluded that word knowledge plays a critical role in the comprehension of written texts. Logically, the relationship between word knowledge and reading comprehension makes sense; in order to understand a passage of text, the reader must know the meanings of many of the words within the text and possess strategies for uncovering the meanings of unknown words.

Although a relationship between word knowledge and reading comprehension has been well documented, the nature of this relationship remains unclear. Anderson and Freebody (1981) generated three hypotheses about the causal relationship between word knowledge and reading comprehension. According to the instrumentalist hypothesis, there is a causal link between vocabulary size and ability to comprehend text. Specifically, having knowledge of more words makes the individual a better reader. In contrast, the knowledge hypothesis suggests that an individual’s background knowledge plays an important role in the comprehension of text. According to this hypothesis, there is a causal link between general knowledge and reading comprehension, and word knowledge is merely one component of general knowledge. Finally, the verbal aptitude hypothesis contends that a third variable, verbal aptitude, explains the relationship between word knowledge and reading comprehension. Specifically, individuals who have high verbal abilities are better word learners and comprehenders than individuals with low verbal abilities. These three conceptualizations of the mechanisms that might underlie the relationship between vocabulary and reading comprehension provide a general context for this study, but the study is not designed to differentiate among them. Rather, the focus of this study is on learning more about potentially different dimensions of vocabulary knowledge and their relationship to reading comprehension.

Evidence from factor analytic studies, correlational studies, experimental studies, and instructional studies supports the relationship between vocabulary knowledge and
reading comprehension. An early factor analytic study (Davis, 1944) found two major factors underlying reading comprehension: word knowledge and reasoning. Davis’ data was later reanalyzed (Spearritt, 1972), and four major factors emerged: word knowledge, drawing inferences from the content, following the structure of the passage, and recognizing a writer’s intent and tone. Although four factors emerged, word knowledge was the best differentiated skill (Spearritt, 1972).

Correlational studies also provide support for the strong relationship between word knowledge and reading comprehension. These studies typically report correlations between vocabulary and reading comprehension that range from .3 to .8. Several factors appear to affect the range of correlations obtained in these studies. For example, measures of reading vocabulary require word reading skills in addition to word knowledge, so they frequently produce higher correlations with reading comprehension than measures of oral vocabulary (cf., Hafner, Weaver, & Powell, 1970; Stanovich & Cunningham, 1992). Additionally, the age of the subjects in the study may affect the correlation between measures of word knowledge and reading comprehension. Several studies have found that the strength of the relationship between word knowledge and reading comprehension increases with age (Snow, 2002; Stanovich, Cunningham, & Feeman, 1984). Finally, the dimension of vocabulary being assessed affects the relationship between word knowledge and reading comprehension. A relatively recent study (Tannenbaum, Torgesen, & Wagner, 2006) found that measures of vocabulary breadth correlated more highly with measures of reading comprehension than measures of vocabulary depth and vocabulary fluency. These three dimensions of word knowledge are an integral part of the current study and will be further discussed later.

Experimental studies provide additional evidence for the relationship between word knowledge and reading comprehension. These studies generally alter the difficulty of the words in a text and examine the impact of word difficulty on comprehension. Replacing words within a text with high frequency synonyms leads to an increase in comprehension, while the introduction of low frequency synonyms inhibits comprehension (Kameenui, Carnine, & Freschi, 1982; Marks, Doctorow, & Wittrock, 1974; Raphael, Myers, Tirre, Fritz, & Freebody, 1981). Word frequency is often used as an indicator of word difficulty because researchers assume that words that are
encountered less frequently in print are less well known. However, word frequency may affect accuracy of word reading in addition to word knowledge, making the results of these studies difficult to interpret.

A final source of evidence for the presence of a relationship between word knowledge and reading comprehension comes from instructional studies that directly teach word meanings to children. Programs that directly teach word meanings to children generally influence knowledge of the taught vocabulary as well as comprehension of texts containing the taught words (Beck & McKeown, 1983; Kameenui et al, 1982; Medo & Ryder, 1993; Stahl & Fairbanks, 1986). It has been more difficult to show that vocabulary instruction improves comprehension of global measures of comprehension such as those found on nationally standardized tests that are not written specifically to contain taught words (Jackson & Dizney, 1963; Tomesen & Aarnoutse, 1998).

Programs that successfully enhance comprehension through vocabulary instruction provide rich instruction and go beyond establishing a single definition for a word (Beck & McKeown, 1983; Beck, McKeown, & Kucan, 2002; Beck, McKeown, & Omanson, 1987). Rich vocabulary instruction involves multiple exposures to the word in varied contexts. This type of instruction develops complex, in-depth knowledge of word meanings as well as knowledge of how words are used in different contexts. Rich vocabulary instruction also enhances the speed of access to word meanings and increases knowledge of semantic relationships (Beck, Perfetti, & McKeown, 1982; Beck et al, 2002). Programs that foster this type of flexible vocabulary knowledge show gains in word knowledge and comprehension of texts with taught words (Beck et al, 1982; McKeown, Beck, Omanson, & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985). Beck and her colleagues have even been successful in improving comprehension of text that does not specifically include taught words through the use of rich vocabulary instruction (Beck et al., 1982).

The techniques involved in developing rich word knowledge through Beck’s instructional methods suggest that vocabulary knowledge is not all-or-none. Specifically, word knowledge exists on a continuum from no knowledge of the word to deep, flexible word knowledge (Beck et al., 2002). This proposition is certainly not new. Many researchers have suggested that different levels or scales of word knowledge exist.
(Anderson & Freebody, 1981; Beck et al., 1982; Cronbach, 1942; Dale, 1965; Stahl, 1985, 1986). Although different labels have been applied to the levels of word knowledge along the continuum, the idea that children can possess different levels of word knowledge is inherent in all of these categorizations. At the most basic level, a child can recognize a word but not provide a definition. As more depth of word knowledge is attained, the child can define the word in greater detail and begin to make associations with other words. Finally, word knowledge becomes decontextualized, and the child develops complex, rich knowledge of the word’s varied meaning in different contexts (Anderson & Freebody, 1981; Beck et al., 1982).

In addition to depth of word knowledge, researchers have also distinguished between vocabulary breadth and vocabulary fluency. Vocabulary breadth refers to the number of words that have some level of meaning to the individual (Anderson & Freebody, 1981). Vocabulary breadth is concerned with the size of the mental lexicon rather than the richness of knowledge that the individual possesses about known words. Vocabulary fluency refers to the rate at which the individual accesses the meaning of the word. The time it takes to access a word meaning decreases as the word meaning becomes reinforced through frequent exposures (Beck et al., 1982; Beck et al., 1987; Wolf, Miller, & Donnelly, 2000). Vocabulary fluency may be one important factor contributing to outcome differences in instructional studies (Beck et al., 2002). Specifically, instructional programs that do not enhance vocabulary fluency may not impact reading comprehension (cf., Jenkins, Pany, & Schreck, 1978).

The present study is a follow-up to a previous study (Tannenbaum et al., 2006) that examined vocabulary breadth, depth, and fluency in their relationships to reading comprehension in third-grade children. Results of confirmatory factor analyses (CFAs) and structural equation modeling (SEM) showed that the three dimensions of word knowledge were not completely distinguishable from one another using the tests employed in the study. Specifically, two dimensions emerged from the analyses: Vocabulary Breadth and Vocabulary Depth/Fluency. Although the two factors of word knowledge both correlated highly with reading comprehension, breadth of word knowledge had a stronger relationship to reading comprehension than did depth/fluency.
However, the two dimensions were highly correlated with one another and shared significant overlapping variance with reading comprehension.

One possible reason that vocabulary depth and vocabulary fluency did not emerge as separate dimensions of word knowledge in the previous investigation is that they may, in fact, be empirically indistinguishable from one another. According to this hypothesis, vocabulary depth and vocabulary fluency are affected by similar types of experiences with words (Tannenbaum et al, 2006). Developing depth of word knowledge involves frequent exposures to a word across different contexts. As flexibility of word knowledge is acquired through multiple encounters with the word, speed of access to the word meaning is also enhanced through the same multiple encounters with the word and perhaps through multiple access routes as well (Beck et al, 2002).

Another possible explanation for the finding of two factors of word knowledge (Tannenbaum et al, 2006) relates to construct validity and the measures employed in the study. According to this hypothesis, vocabulary depth and fluency were not adequately measured as separate constructs with the measures utilized in the study. The authors of the earlier study found some support for this argument through the use of modification indices within the CFA. The specific finding was that model fit improved with the addition of paths between two measures within the depth/fluency dimension and reading comprehension. Of these two measures, one was hypothesized to tap vocabulary depth, while the other was hypothesized to measure vocabulary fluency. Tannenbaum et al (2006) suggested that these two measures of word knowledge may have been assessing other constructs that are more related to reading comprehension than the proposed dimensions of word knowledge. These results also suggest that the two measures were assessing different underlying abilities than their companion measures (Tannenbaum et al, 2006).

In addition to the possible construct validity issue, several of the measures utilized in the Tannenbaum et al study (2006) had relatively low reliabilities within the sample. One advantage to utilizing CFA and SEM is that measurement error is taken into account in the analyses; however, estimates of latent variables are more precise when psychometrically sound measures are used (Kline, 2005). Moreover, one of the measures used by Tannenbaum and her colleagues to measure vocabulary fluency was an
experimenter-developed measure. Although the measure was developed with the conceptualization of vocabulary fluency in mind, limited information about the psychometric properties was available.

Another potential concern with the Tannenbaum et al study (2006) is that the measures of vocabulary fluency could have been measuring general processing speed, making it difficult to determine if general speed was the reason that the measures of vocabulary fluency were related to reading comprehension. However, the finding that depth of word knowledge and fluency of word knowledge held together as one factor despite the nonspeeded nature of the measures of vocabulary depth provides some evidence against this argument. The role of general processing speed will be directly examined within the SEM analyses in this study.

Given the possible issues with construct validity and concerns regarding the psychometric properties of several of the measures utilized in the Tannenbaum et al study (2006), the present study will re-examine the three dimensions of word knowledge in their relationships to reading comprehension using some alternate measures. CFA and SEM will be used to (a) evaluate whether a similar factor structure emerges as the Tannenbaum et al study (2006) with different measures of vocabulary breadth, depth, and fluency in third-grade children, (b) evaluate the factor structure in a sample of seventh-grade children (c) examine the strength of the relationships between the proposed dimensions of word knowledge in both samples, (d) examine the relationships between the dimensions of word knowledge and reading comprehension in both samples, (e) test for measurement equivalence across the two samples of children, (f) test for invariant structural parameters across the two samples, and (g) test for differences in latent means across the two samples.

One of the main goals of the present study is to replicate the findings of the Tannenbaum et al study (2006) with a similar sample of third-grade children using different measures of vocabulary breadth, depth, and fluency. Specifically, this study will address whether the factor structure of word knowledge depends on the measures employed in the study. Further, this study will utilize measures that have undergone adequate reliability and validity testing, when possible.
In another expansion from the earlier study (Tannenbaum et al, 2006), the current study will examine the factor structure of word knowledge in a sample of seventh-grade children to evaluate whether the structure of word knowledge or the relationships between word knowledge and reading comprehension change with age. A previous study (Schatzschneider, Harrell, & Buck, 2007) found that vocabulary knowledge, in addition to other measures of reasoning ability, accounted for more unique variance in reading comprehension in a sample of seventh-grade children than a sample of third-grade children.

As children advance from elementary school to middle school, vocabulary demands become greater because content area reading becomes more prominent (Harmon, Hedrick, & Wood, 2005). Within content area reading, many of the vocabulary words are low frequency words that are content specific (Harmon, Hedrick, Wood, & Gress, 2005). Moreover, many of the content words children encounter on a daily basis are new labels for underlying concepts that may or may not already be known (Alvermann & Phelps, 2002; Harmon, Hedrick, & Wood, 2005). Additionally, many of the words middle school children encounter have different meanings from the ones they are familiar with. Middle school children may possess knowledge of a common meaning of a word, but when that word is encountered in a content area text, the meaning of the familiar word changes. In other words, known labels can have specialized meanings when encountered in a content area text (Alvermann & Phelps, 2002; Harmon, Hedrick, & Wood, 2005).

Knowledge of multiple meanings of words may become more important in middle school because words within content area texts can be multidimensional (Beck et al, 2002). The vocabulary load present in content area texts implies that middle school children need to possess a broad range of word knowledge and thorough understanding of known words. Therefore, both breadth and depth of word knowledge may be particularly important components of vocabulary knowledge necessary for comprehension of written texts in middle school.

Given that the current study will utilize two samples of different aged children, other goals of the study will be to examine measurement equivalence, invariance of structural parameters, and differences in latent means. An assumption of cross-sectional
designs in developmental research is that the eventual behavior of a younger group of individuals can be estimated from the behavior of a different, older group of individuals (Achenbach, 1978). Cross-sectional designs can provide information about age group differences (Miller, 1998); however, a major methodological concern of cross-sectional designs is whether the measurement instruments assess similar things at different ages (Knight & Hill, 1998; Labouvie, 1980; Miller, 1998; Nunnally, 1973). The concern of measurement equivalence is essentially an issue of construct validity, and several statistical approaches can be utilized to assess measurement invariance. Multiple-group CFA is the primary approach to assessing measurement invariance in developmental research designs (Byrne, Shavelson, & Marsh, 1992; Byrne, Shavelson, & Muthen, 1989; Hartmann, 2005; Meade & Lautenschlager, 2004). Within multiple-group CFA, parameters can be tested for invariance across groups through a series of increasingly restrictive tests (Byrne, 2001). Generally, the measurement parameters are examined first through tests of invariance of factor loadings. Structural parameters, such as factor variances/covariances and structural regression paths, are then examined for structural invariance across groups (Byrne, 2001; Byrne et al, 1989). Other statistical methods have also been suggested for assessing measurement invariance, such as comparison of reliability coefficients across age groups and comparison of correlations with other measures across age groups (Knight & Hill, 1998). Means of latent variables can also be estimated within CFA and SEM to test for group differences in the means of the latent constructs (Byrne, 2001; Kline, 2005).

There are several important theoretical and practical implications of this study. First, this study will attempt to validate the findings of the Tannenbaum et al study (2006) to determine if the three proposed dimensions of word knowledge are distinguishable from one another in third-grade children. If similar results are found with different measures of word knowledge, this study will provide more support for a two-factor model of word knowledge in third-grade children. Second, this study will examine whether there are developmental differences in the structure of word knowledge by also utilizing a sample of seventh-grade children. If a different structure of word knowledge emerges, this will be the first study to empirically demonstrate that the structure of word knowledge may become more or less differentiated with age. Third, this study will
examine the relationships between word knowledge and reading comprehension across both grades. If there are reliable differences in the relationships between the different dimensions of word knowledge and reading comprehension across grades, this may have important implications for different instructional emphases at these two grade levels.
METHOD

Participants

Data for this study were collected from 265 third-grade students and 166 seventh-grade students from a suburban school district in Tallahassee, Florida. Students were drawn from five elementary schools and three middle schools representing diverse socioeconomic backgrounds. Consent forms were sent home with all children in the classrooms of all teachers who agreed to participate in the study. All students who returned consent forms were tested. One third-grade student’s data was not included because English was not his primary language. Data from one seventh-grade student was excluded because he was severely hearing impaired. The final samples contained 264 third-grade students and 165 seventh-grade students.

The third-grade sample consisted of 140 girls (53%) and 124 boys (47%), and the seventh-grade sample consisted of 99 girls (60%) and 66 boys (40%). The ethnic composition of the third-grade sample was comprised of 45.1% African American, 49.2% Caucasian, 2.3% Asian, 1.9% Hispanic, and 1.5% Other. The ethnic composition of the seventh-grade sample was comprised of 39.4% African American, 54.5% Caucasian, 1.8% Asian, 3.0% Hispanic, and 1.2% Other. The third-grade sample ranged in age from 8 years, six months to 11 years, five months ($M = 8.84$ years, $SD = .64$), and the seventh-grade sample ranged in age from 11 years, 10 months to 15 years, two months ($M = 12.99$ years, $SD = .69$). Approximately 39.4% of the third-grade sample and 36.4% of the seventh-grade sample were eligible for free or reduced priced lunch.

To examine the comparability of the two samples to the populations from which they were drawn, scores on the FCAT-SSS in the third- and seventh-grade samples were compared to the average scores for each school participating in this study, as well as the average scores for third- and seventh-grade children in the Leon County School District (see Tables 1-2). In general, the participants in this study scored slightly higher on the FCAT-SSS than the average scores at the district and school level, suggesting that the two samples in this study had slightly higher reading comprehension scores than the populations of students from which they came.
**Vocabulary Measures**

Two measures from each of the three proposed dimensions of word knowledge were administered to each participant. The Expressive One-Word Picture Vocabulary Test, Third Edition (Brownell, 2000) and the Receptive One-Word Picture Vocabulary Test, Second Edition (Brownell, 2000) were used to assess vocabulary breadth. The Multiple Contexts subtest of the Test of Word Knowledge (Wiig & Secord, 1992) and the Associations subtest of The WORD Test-2 (Bowers, Huisingh, LoGiudice, & Orman, 2004, 2005) were used to measure vocabulary depth. The Picture Naming: Nouns subtest of the Test of Word Finding-2/Test of Adolescent and Adult Word Finding (German, 1990, 2000) and the Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities (McGrew & Woodcock, 2001) were used to assess vocabulary fluency.

**Expressive One-Word Picture Vocabulary Test, Third Edition (EOWPVT).** The EOWPVT (Brownell, 2000) is a measure of expressive vocabulary that can be used with individuals aged two through 18 years. The participant is shown a series of pictures and asked to provide the name of the object, action, or concept depicted in each picture. The items become progressively more difficult as the test continues. Testing is discontinued when the participant makes six consecutive errors. Scores can be represented as standard scores, percentile ranks, and age equivalents. During development of the test, items were selected from a variety of sources to represent words that are expected to be known at different ages. Items were obtained from parent questionnaires of words spoken by children at different ages, word frequency guides, and grade-level curriculum materials. The final form of the third edition of the EOWPVT represents a range of word knowledge from familiar to obscure for all age levels assessed. Brownell (2000) reports a median internal consistency reliability of .96, a median split-half reliability of .98, a test-retest reliability of .90, and an inter-rater reliability of .99. Criterion-related validity was established by examining correlations between the EOWPVT and several other expressive and receptive vocabulary tests. These correlations range from .67-.90, with the EOWPVT correlating highest with other measures of expressive vocabulary (e.g., Vocabulary subtests of the Wechsler Intelligence Scale for Children-III and Stanford Binet-4). A relationship was also established between scores on the EOWPVT and
general cognitive ability. A correlation of .89 was found between the EOWPVT and the Otis-Lennon School Ability Test-Seventh Edition (Brownell, 2000). Internal consistency reliability of the EOWVPT in the third- and seventh-grade samples in this study was .95 and .94, respectively.

**Receptive One-Word Picture Vocabulary Test, Second Edition (ROWPVT).** The ROWPVT (Brownell, 2000) is the companion measure to the EOWPVT and assesses receptive vocabulary. It can be used with individuals aged two through 18 years. The participant selects an illustration that best depicts the meaning of a word spoken by the examiner. The items become progressively more difficult, and testing is discontinued when the participant makes six errors out of eight consecutively administered items. Scores can be reported as standard scores, percentile ranks, and age equivalents. Similar to the EOWPVT, items for the ROWPVT were selected to represent a range of word knowledge for all ages. Items were selected based on word frequency, difficulty, grade level at which the words appear in curriculum materials, and teacher report of appropriateness. Brownell (2000) reports a median internal consistency reliability of .96, a median split-half reliability of .98, a test-retest reliability of .84, and an inter-rater reliability of 1.00. Criterion-related validity was established by correlating the ROWPVT with other tests of expressive and receptive vocabulary. These correlations ranged from .44-.97, with a median correlation of .71. The highest correlations were found between the ROWPVT and measures of expressive vocabulary that assess a broad range of word knowledge (e.g., Vocabulary subtests of the Wechsler Intelligence Scale for Children-III and Stanford Binet-4). Brownell (2000) also established a strong relationship between the ROWPVT and the Otis-Lennon School Ability Test-Seventh Edition, a measure of general cognitive ability (r = .77). Internal consistency reliability of the ROWPVT in the third- and seventh-grade samples in this study was .94 and .95, respectively.

**Multiple Contexts subtest of the Test of Word Knowledge (TOWK).** The Multiple Contexts subtest of the TOWK (Wiig & Secord, 1992) requires the participant to describe two contexts or meanings for each stimulus word. The stimulus words are presented to the participant orally and visually. A correct response refers to two independent contexts or meanings. For example, the participant is presented with the word *ship* and must indicate two separate meanings for the stimulus word (i.e. *to send something far away*
The number of different meanings for the stimulus words ranges from two to five. Testing is discontinued when the participant receives five consecutive scores of zero. The Multiple Contexts subtest can be used with individuals aged eight to 17 years. Scores are reported as subtest standard scores. Items for the Multiple Contexts subtest were selected based on word frequency, difficulty, grade level at which the word is recognized, and having multiple definitions (Wiig and Secord, 1992). The authors of the test report internal consistency reliabilities and test-retest reliabilities of the Multiple Contexts subtest that ranges from .87-.92 and .85-.90, respectively, based on age. The developers of the test examined the validity of the TOWK by evaluating the extent to which the test correctly identified participants as language-learning disabled (LLD) or non-LLD, as previously established by each participant’s school record. The overall test correctly identified LLD and non-LLD 67 percent of the time. Mean scores of the LLD and non-LLD participants on the Multiple Contexts subtest were 6.9 and 10.1, respectively. Although a direct relationship between the Multiple Contexts subtest and cognitive ability was not assessed, Wiig and Secord (1992) report a correlation of .73 between the expressive composite of the TOWK and the WISC-III. Internal consistency reliability of the Multiple Contexts subtest in the third- and seventh-grade samples in this study was .85 and .90, respectively.

**Associations subtest of The WORD Test-2.** The Associations subtest of The WORD Test-2 (Bowers, Huisingh, LoGiudice, & Orman, 2004, 2005) assesses the ability to build semantic relationships among words by requiring the participant to make associations among words and link the words to a common dimension. The participant is presented with four words and must choose the one semantically unrelated word. The participant then explains the choice in relation to the other words. For example, the participant is presented with the words *plot, setting, scissors,* and *climax* and must correctly indicate that the word *scissors* does not belong because the other words represent parts of literature. The participant must also have knowledge of the multiple meanings of a word like *climax* or *setting,* and must be able to use them flexibly in order to link them to the word *plot,* as parts of literature. Two forms of this test exist, one form for elementary school children and another form for adolescents. The participant is administered all fifteen items. Standard scores, percentile ranks, and age equivalents can
be reported for performance on this subtest. Items were selected to include vocabulary from grade-level curricula. For the Elementary form, the authors report an average internal consistency reliability of .74 and an average test-retest reliability of .86. The average internal consistency reliability and average test-retest reliability reported by the authors for the Adolescent form are .76 and .71, respectively (Bowers et al, 2004, 2005). Validity was established by examining differences in mean scores between participants in the normative sample and a matched sample of language-disordered participants. The authors found that the language-disordered sample performed significantly lower than the participants in the normative sample (Bowers et al, 2004, 2005). The relationship between the Associations subtest and cognitive ability was not examined by the authors. Internal consistency reliability of the Associations subtest in the third- and seventh-grade samples in this study was .72 and .61, respectively.

Picture Naming: Nouns subtest of the Test of Word Finding-2 (TWF-2) and Test of Adolescent/Adult Word Finding (TAWF). The Picture Naming: Nouns subtest of the TWF-2 (German, 2000) and TAWF (German, 1990) assesses accuracy and speed of naming pictures of objects. The participant is shown a picture of an object and must provide the correct target item name. Two forms of the test exit: the TWF-2 for ages four years to 12 years and the TAWF for ages 12 years to 80 years. The Picture Naming: Nouns subtest on the TWF-2 contains 22 items, and the version on the TAWF contains 37 items. Items were selected for the TWF-2 and TAWF based on their semantic relations, syntactic features, phonological features, comprehensibility, and word frequency (German, 1990, 2000). Criterion-related validity was established for both tests by correlating them with other confrontation naming tests. These correlations ranged from .53-.69 for the TWF-2 and .62-.66 for the TAWF (German, 1990, 2000). Although standard scores can be derived on both forms of the test, the TAWF standard scores do not take into account speed of response. In order to ensure that this task measures vocabulary fluency, the administration procedures were slightly modified. Specifically, response time latencies were digitally recorded and measured using the Audacity software program. Specifically, participants’ audio responses were recorded on a computer, and response time latencies were manually measured for every item. Participants were administered all items on the appropriate form for their age. Pilot
testing was conducted with samples of third- (n = 25) and seventh-grade (n = 19) children in the fall of 2006 to verify that the reliabilities of the tests remain high under the new administration conditions. Reliability was established by examining the frequencies of every response time within both samples and determining the time at which 60 percent of responses were correct. This procedure led to cutoff times of 1.25 seconds and .9 seconds for the third- and seventh-grade samples, respectively. These cutoff times were used to establish scores of 1 or 0 for responses on all items. Total scores were then calculated for each participant. Using this procedure, internal consistency reliability of the Picture Naming: Nouns subtest was .82 and .87 for the third- and seventh-grade pilot samples, respectively. These values are comparable to the internal consistency reliabilities reported by German (1990, 2000) for the standard administration and scoring procedures. Using these same procedures, internal consistency reliability for the third- and seventh-grade samples in this study was .74 and .86, respectively. Since response times were manually measured from each participant’s Audacity sound file, inter-rater reliability was calculated. Response times were measured by two raters for 30 percent of the cases in both samples. Total scores were then calculated for these participants, and a correlation was obtained between the total scores for both raters. The two raters were the author and an undergraduate assistant. Inter-rater reliability in the third-grade and seventh-grade samples was .98 and .99, respectively.

*Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities.* The Retrieval Fluency subtest (McGrew & Woodcock, 2001) requires the participant to name as many items as possible from three separate categories, with one minute allowed for responding to each category. The categories administered are animals, names, and food or drinks. This test can be used with individuals aged two to 80+ years. Scores can be reported as standard scores, percentile ranks, age equivalents, and grade equivalents. Items were selected for the test to represent a measure of ideational fluency. Given the speeded nature of the test, McGrew and Woodcock (2001) used Rasch analyses to determine the reliability of the test. They report a median reliability of .85. Validity was established by utilizing CFA to demonstrate that this narrow measure of ideational fluency mapped onto the broader factor of Long-Term Retrieval. The authors also examined the correlations between Long-Term Retrieval and several measures of
cognitive ability. The correlations ranged from .52-.69 (McGrew & Woodcock, 2001). Rasch analyses were conducted through the Winsteps software program (Linacre, 2007) to examine the reliability of the Retrieval Fluency subtest in both samples in this study. Reliability in the third- and seventh-grade samples was .68 and .79, respectively.

**Processing Speed Measures**

**Decision Speed subtest of the Woodcock Johnson III Tests of Cognitive Abilities.** The Decision Speed subtest (McGrew & Woodcock, 2001) requires the participant to look at rows of pictures and select the two pictures within each row that are related as quickly as possible. The participant is given three minutes to complete as many of the 40 items as possible. The Decision Speed subtest can be used with individuals aged two to 80+ years. Scores can be reported as standard scores, percentile ranks, age equivalents, and grade equivalents. Items were selected for the test to represent a measure of semantic processing speed. Given the speeded nature of the test, McGrew and Woodcock (2001) used Rasch analyses to determine the reliability of the test. They report a median reliability of .87. Validity was established by utilizing CFA to demonstrate that this narrow measure of semantic processing speed mapped onto the broader factor of Processing Speed. The authors also examined the correlations between the Processing Speed factor and several measures of cognitive ability. The correlations ranged from .43-.56 (McGrew & Woodcock, 2001). Rasch analyses indicated that the reliability of the Decision Speed subtest in the third- and seventh-grade samples in this study was .90 and .77, respectively.

**Visual Matching subtest of Woodcock Johnson III Tests of Cognitive Abilities.** The Visual Matching subtest (McGrew & Woodcock, 2001) requires the participant to look at rows of numbers and select the two numbers that are alike as quickly as possible. The test contains single-digit, two-digit, and three-digit numbers. The participant is given three minutes to complete as many of the 60 items as possible. This test can be used with individuals aged two to 80+ years. Scores can be reported as standard scores, percentile ranks, age equivalents, and grade equivalents. Items were selected for the test to represent a measure of perceptual speed. The Visual Matching subtest and Decision Speed subtest are companion measures designed to assess general processing speed. These two tests were designed to measure qualitatively different narrow abilities that both load on the
same broad factor of Processing Speed. Given the speeded nature of the test, McGrew and Woodcock (2001) used Rasch analyses to determine the reliability of the test. They report a median reliability of .91. Validity was established by utilizing CFA to demonstrate that this narrow measure of perceptual speed loaded onto the broader factor of Processing Speed. As previously stated, the authors found correlations between the Processing Speed factor and cognitive ability that ranged from .43-.56 (McGrew & Woodcock, 2001). Rasch analyses indicated that the reliability of the Visual Matching subtest in the third- and seventh-grade samples in this study was .96 and .94, respectively.

**Reading Comprehension Measures**

The reading portions of the Florida Comprehensive Assessment Test-Sunshine State Standards (FCAT-SSS) and the Stanford Achievement Test-Tenth Edition (SAT-10; Harcourt Assessment, 2003) were used to assess reading comprehension. The FCAT-SSS and SAT-10 are standardized tests of achievement that are administered to public school students in grades 3 to 10 in the state of Florida every spring. Scores from the reading portions of these two tests were obtained from the school district for each participant. Although both tests require students to read passages and then answer multiple-choice questions, the passages are constructed somewhat differently on the FCAT-SSS and SAT-10. Specifically, the FCAT-SSS passages are generally chosen from previously published writings, whereas the SAT-10 passages are constructed for the test. Further, the quantity of literary and informational passages is equal on the SAT-10, whereas the FCAT-SSS contains more literary passages. Finally, the FCAT-SSS passages tend to be longer than the SAT-10 passages (Florida Department of Education, 2000, 2002, 2005). Although differences exist in the construction of the FCAT-SSS and the SAT-10, questions on the two tests measure similar reading comprehension skills. For example, both tests assess the ability to understand explicitly stated facts, understand implied ideas and messages, classify and sequence information, compare and contrast information, and determine cause and effect (Florida Department of Education, 2005, 2006).

**FCAT-SSS Reading.** The FCAT-SSS is a criterion-referenced, standardized test that measures achievement by determining if students are meeting the Sunshine State Standards. The Standards are statements of skills and competencies that students should
possess at each grade level. These competencies are called “benchmarks” and are expected to be taught by the classroom teacher (Florida Department of Education, 2002). The questions on the test match the complexity of the standards being assessed, and the questions vary in difficulty in order to capture the proficiency level of all students (Florida Department of Education, 2001). Students in grade 3 must demonstrate proficient reading skills on the FCAT before being promoted to grade 4. Scores are reported on a scale of 100 to 500, with a mean of 300 and standard deviation of 50. Achievement level scores are also reported, which range from 1 (lowest) to 5 (highest) (Florida Department of Education, 2001, 2006). Reading comprehension skills tapped by the FCAT-SSS include identification of the author’s purpose, determination of the stated or implied main idea, identification of relevant facts and details, recognition and arrangement of events in chronological order, and understanding of plot development and conflict resolution (Florida Department of Education, 2006). Reliability of the reading portion of the FCAT-SSS ranges from .85-.90, with a reliability of .89 in the third-grade and .90 in the seventh-grade (Florida Department of Education, 2007).

SAT-10 Reading. The SAT-10 (Harcourt Assessment, 2003) is a norm-referenced, standardized test that allows for the comparison of the performance of Florida public school students to national norms (Florida Department of Education, 2005). The questions vary in difficulty in order to capture the achievement of all students (Florida Department of Education, 2006). Scale scores on the SAT-10 are constant scores that show performance along a single, comparable scale across grade levels. Thus, the SAT-10 scale scores can be used to show how much a student has progressed from year to year (Florida Department of Education, 2006). National Percentile Ranks and Stanine scores can be calculated from a student’s scale score and indicate relative standing compared to students at the same grade nationwide. National Percentile Ranks range from 1 to 99, while Stanine scores range from 1 (lowest) to 9 (highest) (Florida Department of Education, 2006). Reading comprehension skills tapped by the SAT-10 include the ability to comprehend explicitly stated relationships, the ability to form an interpretation based on explicit and implicit information, and the ability to synthesize and evaluate explicit and implicit information (Florida Department of Education, 2006). Reliability of
the reading portion of the SAT-10 ranges from .90 to .92, with a reliability of .92 in the third-grade and .91 in the seventh-grade (Florida Department of Education, 2007).

Procedure

The six measures of word knowledge and two measures of processing speed were administered to participants in the spring of 2007 in order to compare performance on these measures with scores on the FCAT-SSS and SAT-10. All participants were tested after the schools finished the administration of the FCAT-SSS and SAT-10. Each participant was tested individually by the author or a trained undergraduate student. Student testers were trained on the measures of word knowledge for approximately 15 hours, and they were required to demonstrate proficient administration before being allowed to test participants.

Participants were tested across two sessions that each lasted for approximately 30 minutes. One measure from each of the three hypothesized dimensions of word knowledge and one measure of processing speed were administered during each session. During the first testing session, the EOWPVT, the Associations subtest of The WORD Test-2, the Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities, and the Decision Speed subtest of the Woodcock Johnson III Tests of Cognitive Abilities were administered. The second testing session was comprised of the ROWPVT, the Multiple Contexts subtest of the Test of Word Knowledge, the Picture Naming: Nouns subtest of the TWF-2/TAWF, and the Visual Matching subtest of the Woodcock Johnson III Tests of Cognitive Abilities. This format was used in order to limit error variance due to time sampling. Additionally, the schools requested that we limit our testing to half-hour sessions.
Table 1
Comparison of Mean Scores on the FCAT-SSS between Third- and Seventh-Grade Students in the Current Samples and in the School District

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>Current Sample</th>
<th>District</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third-Grade</td>
<td>332</td>
<td>322</td>
</tr>
<tr>
<td>Seventh-Grade</td>
<td>330</td>
<td>321</td>
</tr>
</tbody>
</table>
Table 2
Comparison of School-Level Mean Scores on the FCAT-SSS between Participants in the Current Study and the Entire Grade

<table>
<thead>
<tr>
<th>School Name</th>
<th>Participants</th>
<th>Entire Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hartsfield Elementary</td>
<td>310</td>
<td>300</td>
</tr>
<tr>
<td>Oak Ridge Elementary</td>
<td>274</td>
<td>271</td>
</tr>
<tr>
<td>Ruediger Elementary</td>
<td>313</td>
<td>303</td>
</tr>
<tr>
<td>Springwood Elementary</td>
<td>338</td>
<td>318</td>
</tr>
<tr>
<td>Killearn Lakes Elementary</td>
<td>370</td>
<td>357</td>
</tr>
<tr>
<td>Seventh Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BelleVue Middle School</td>
<td>296</td>
<td>276</td>
</tr>
<tr>
<td>Cobb Middle School</td>
<td>343</td>
<td>328</td>
</tr>
<tr>
<td>Swift Creek Middle School</td>
<td>337</td>
<td>326</td>
</tr>
</tbody>
</table>
RESULTS

Data Issues and Descriptive Statistics

All analyses were conducted with unstandardized variables, when possible. Although there is a general preference for using standardized variables in research, the use of standardized variables in SEM can be problematic for several reasons. First, most estimation methods in SEM assume the use of unstandardized variables and may yield incorrect results when used with standardized variables. Second, important information may be lost when standardized variables are used in SEM analyses involving multiple samples (Kline, 2005). The use of unstandardized variables was not possible for the measurement of reading comprehension because raw scores are not reported for these measures.

The third-grade and seventh-grade datasets were examined for missing data, univariate outliers, bivariate outliers, multivariate outliers, skewness, and kurtosis. There were seven missing data points in the third-grade sample (four for FCAT-SSS and three for SAT-10) and two missing data points in the seventh-grade sample (both for SAT-10). Full-information maximum-likelihood estimation (FIML) was used through the AMOS software program (SPSS, 2007) to handle the missing data points. This approach was utilized because maximum likelihood estimates of missing data provide the least biased estimates (Byrne, 2001; Little & Rubin, 1989). Twenty-two univariate outliers (seven for Associations, five for Retrieval Fluency, three for Decision Speed, five for ROWPVT, and two for FCAT-SSS) were identified in the third-grade data set and 13 univariate outliers (three for Associations, one for Decision Speed, two for Visual Matching, three for FCAT-SSS, and four for SAT-10) were identified in the seventh-grade data set. These values were recoded to values that were two interquartile ranges from the median. Visual inspection of scatter plots revealed no bivariate outliers. To determine if multivariate outliers were present, cases were sorted by Mahalanobis distance. A conservative probability estimate, \( p < .001 \), was used to identify outliers (Tabachnick & Fidell, 2001). This procedure revealed no multivariate outliers. The resulting skewness and kurtosis values fell within acceptable ranges.
Means, standard deviations, reliabilities, and intercorrelations for both samples are reported in Tables 3 and 4. Of note, the two measures of processing speed were not strongly related to the majority of vocabulary measures in both samples. However, the two measures of processing speed were both related to Retrieval Fluency, a measure of vocabulary fluency, providing some evidence that one of the measures of vocabulary fluency may be measuring general processing speed. Further, the scores on Retrieval Fluency in both samples were not highly correlated with any of the measures of vocabulary breadth or depth.

Examination of the correlation between the two measures of vocabulary fluency in both samples indicates that the relationship between these two measures is weaker than the relationships between the two companion measures of breadth and depth. Ninety-five percent confidence intervals were created for the correlations between the companion measures of breadth, depth, and fluency of word knowledge. Examination of these confidence intervals for overlap further demonstrates that the correlation between the two vocabulary fluency measures in both samples was statistically weaker than the correlations between the two measures of the other dimensions of word knowledge. For example, the correlation of .22 between Retrieval Fluency and the Test of Word Finding in the third-grade sample was significantly lower than the corresponding correlations between the two measures of vocabulary depth and the two measures of vocabulary breadth. At seventh grade, the correlation of .32 between Retrieval Fluency and the Test of Adolescent Word Finding was significantly smaller than the correlation of .74 between the two measures of vocabulary breadth, and the correlation of .58 between the measures of vocabulary depth.

Evaluation of Measurement Models in the Third-Grade Sample

Confirmatory factor analyses were conducted using Amos 16.0 (SPSS, 2007) to evaluate measurement models. All analyses were based on covariance matrices. A model that represented vocabulary Breadth, Depth, and Fluency as three separate latent variables was compared to nested models to determine if any two-factor or one-factor models provided a better fit to the data. All models were identified, indicating that a unique set of parameter estimates were obtained. Fit indices for the different models are presented in Table 5. The p-value associated with the $\chi^2$ test is the probability of obtaining
the observed variance/covariance matrix if the model is true for the population. Thus, a nonsignificant $\chi^2$ is desired. Values less than 3.0 are desirable for the $\chi^2$/df test. Following the suggestions of Hu and Bentler (1998), a cutoff value of .95 was used as an indicator of good fit for the TLI and CFI indices, and a cutoff of .06 was used for the RMSEA index. Values greater than .9 for the AGFI index were used as an indication of good fit (Bentler & Bonett, 1980).

Examination of the models indicates that the solutions for the three-factor model, the two-factor model of Fluency and Breath/Depth, and the two-factor model of Breath and Depth/Fluency were not admissible due to multicollinearity. These three models are presented in Figures 1-3 with standardized regression weights and error variances. The solutions for the two-factor model of Depth and Breadth/Fluency and the one-factor model of General Word Knowledge were admissible. These models are presented in Figures 4-5. The majority of the fit indices suggest that these two models provided a good fit to the data; however, the results of the $\chi^2$ tests indicate a low probability of obtaining the observed variance/covariance matrices if the models are true for the population. Comparison of these two models with a $\chi^2$ difference test revealed that the one-factor model provided the best fit to the data. A very high correlation was found between Depth and Breadth/Fluency in the two-factor model, also indicating that a one-factor model may be more appropriate. All further analyses for the third-grade sample were conducted with the one-factor model of General Word Knowledge.

Given the finding of a low p-value associated with the $\chi^2$ test for the one-factor model, misspecification was examined through the use of modification indices. Results suggested that model fit would improve with the addition of an error covariance between the EOWPVT and the TWF. The addition of this covariance makes substantive sense, as both tests utilize a similar method of requiring the participant to name an object depicted in a picture. Fit indices for this model are presented in Table 5 and suggest a good fitting model. All further analyses for the third-grade sample were conducted with the one-factor model of General Word Knowledge with an error covariance specified between the EOWPVT and the TWF.
Evaluation of Measurement Models in the Seventh-Grade Sample

CFAs for the seventh-grade sample were based on covariance matrices. All models were identified. Fit indices for the different models are presented in Table 6. Taken together, these indices indicate that none of the models provided a good fit to the data. The different models are presented in Figures 6-10 with standardized regression weights and error variances. Examination of the correlations between the different factors of word knowledge revealed that Breadth and Depth were highly correlated, suggesting that they might be better represented as one factor. Results of $\chi^2$ difference tests provide additional evidence that a two-factor model of Fluency and Breadth/Depth was the best fitting model. Further, the standardized regression weights for the two indicators of vocabulary fluency were highest when fluency was represented as a separate factor. Overall, results suggest that the two-factor model of Fluency and Breadth/Depth provided the best fit to the data in the seventh-grade sample.

Although a two-factor model of Fluency and Breadth/Depth emerged as the best fitting model, the fit indices suggest poor overall model fit. Modification indices were also used in the seventh-grade sample to examine misspecification. Similar to the third-grade model, the modification indices revealed that overall model fit would improve with an error covariance between the EOWPVT and the TAWF. Fit indices for this model are presented in Table 6 and suggest a good fitting model. All further analyses for the seventh-grade sample were conducted with a two-factor model of Fluency and Breadth/Depth with an error covariance specified between the EOWPVT and the TAWF.

Evaluation of Structural Models in the Third-Grade Sample

Structural equation modeling was used to examine the predictive relation between the factors of General Word Knowledge and Reading Comprehension in the third-grade sample. Structural equation modeling allows for the evaluation of the contribution of each predictor variable to the outcome variable with all other variables held constant. The structural portion of the one-factor model in the third-grade sample is presented in Figure 11. Fit indices for the model suggest good overall fit (see Table 5). Seventy-six percent of the variance in Reading Comprehension was explained by the General Word Knowledge factor. Analysis of the structural regression weight indicates that Reading Comprehension increased .87 for each unit increase in General Word Knowledge.
The Decision Speed and Visual Matching subtests were added as a Processing Speed factor to the model to examine the effects of Processing Speed as a predictor variable. This model is presented in Figure 12. Results of a $\chi^2$ difference test indicated that Processing Speed and General Word Knowledge were distinct dimensions ($\chi^2$ diff = 84.79, df = 1, $p < .001$). The addition of Processing Speed had minimal impact on the structural coefficient for General Word Knowledge and the amount of variance in Reading Comprehension accounted for by the independent variables.

Analyses of the relationship between general word knowledge and comprehension were also conducted with FCAT-SSS and SAT-10 as individual dependent variables to determine if similar results would be obtained for both measures of Reading Comprehension. These models are presented in Figures 13-14 with standardized regression weights and error variances. Examination of the regression weights and error variances indicates that the FCAT-SSS and SAT-10 performed comparably in the two models.

**Evaluation of Structural Models in the Seventh-Grade Sample**

The structural portion of the two-factor model of Fluency and Breadth/Depth in the seventh-grade sample is presented in Figure 15. Fit indices for the model suggest adequate overall fit (see Table 6). Sixty-four percent of the variance in Reading Comprehension was explained by the Fluency and Breadth/Depth factors combined. Analysis of the structural regression weights indicates that Reading Comprehension increased .75 for each unit increase in Breadth/Depth. The path between Fluency and Reading Comprehension was not significant when it was included in the model with Breadth/Depth. Breadth/Depth accounted uniquely for 28% of the variance in Reading Comprehension. Thus, 36% of the variance in Reading Comprehension was accounted for by constructs jointly measured by fluency and breadth/depth.

The relationships between the two factors of word knowledge and Reading Comprehension in the seventh-grade sample were also examined in a CFA to allow for the inspection of the simple correlations between the latent variables. Results of the CFA reveal correlations of .80 and .52 between Reading Comprehension and Breadth/Depth and Fluency, respectively. Examination of only the regression weights in the structural model may lead the reader to think that Fluency is not contributing to the prediction of
Reading Comprehension. Taken together, the results of the CFA and SEM indicate that Fluency is moderately related to Reading Comprehension and makes its contribution to the prediction of Reading Comprehension through overlapping variance with Breadth/Depth.

The effects of Processing Speed as an independent variable were also examined in the seventh-grade structural model. The model is presented in Figure 16. A moderately-sized correlation was found between vocabulary Fluency and Processing Speed; however, results of a $\chi^2$ difference test indicated that the two dimensions were distinct ($\chi^2 \text{ diff} = 20.8$, df = 2, $p < .001$). Processing Speed contributed to the overall variance accounted for in the Reading Comprehension factor. Thus, for seventh graders, it appears that the processing speed tasks assess some capability that is also assessed on the reading comprehension test, but which is not assessed by the vocabulary measures.

The addition of a Processing Speed factor to the SEM analysis also allowed for examination of the role of general processing speed in the predictive relation between Vocabulary Fluency and Reading Comprehension. If general processing speed is one reason that vocabulary fluency is related to reading comprehension, there should be a decrease in the structural coefficient for Vocabulary Fluency when Processing Speed is added to the model. Since the structural coefficient for Vocabulary Fluency was not significantly different from zero in the SEM analysis, additional SEM analyses were conducted with only the Processing Speed and Vocabulary Fluency factors in the model. Comparison of a model that included both factors and a model that included only Vocabulary Fluency show that the structural coefficient for Vocabulary Fluency decreased with the addition of the Processing Speed factor (see Figures 17 and 18); however, the reduction in the structural coefficient was not statistically reliable. Therefore, the difference in the structural coefficients could be due to random effects in the models.

The FCAT-SSS and SAT-10 were also examined as individual dependent variables in the seventh-grade sample. These models are presented in Figures 19-20 with standardized regression weights and error variances. Examination of the regression weights and error variances indicates that the FCAT-SSS and SAT-10 performed comparably in the two models.
Evaluation of Measurement Invariance

The finding of unequal numbers of factors in the third-grade and seventh-grade samples suggests that the measures of word knowledge were measuring different constructs in the two samples, or that the constructs themselves are differentially related to one another in third- and seventh-grade students. Testing for measurement equivalence does not necessarily require an identical number of factors across the two groups; however, only the comparable parameters within each factor can be equated (Werts, Rock, Linn, & Joreskog, 1976). Therefore, only the indicators of Reading Comprehension could be tested for invariance within a CFA model. A model constraining the loading of the FCAT-SSS to be equivalent across both groups ($\chi^2 = 44.2, \text{df} = 35$) was compared to a model that allowed the loading to be freely estimated in both groups ($\chi^2 = 41.8, \text{df} = 34$). Results of the $\chi^2$ difference test indicate that the FCAT-SSS was invariant across the two groups ($\chi^2\text{ diff} = 2.4, \text{df} = 1, p > .10$). Similarly, a model constraining the loading of the SAT-10 to be equivalent across both groups ($\chi^2 = 44.2, \text{df} = 35$) was compared to a model that allowed the loading to be freely estimated in both groups ($\chi^2 = 41.8, \text{df} = 34$). Results of the $\chi^2$ difference test indicate that the SAT-10 was also invariant across the two groups ($\chi^2\text{ diff} = 2.4, \text{df} = 1, p > .10$). Thus, the measures of reading comprehension were related identically to the unobserved latent variable across both samples.

Although the indicators of word knowledge did not load onto the same factors in the two samples, invariance of the loadings was examined through the use of logically ordered models with equality constraints imposed on certain parameters (Byrne, 2001). First, a baseline model was run in which all parameters were freely estimated for both groups ($\chi^2 = 41.8, \text{df} = 34$). This model provided the $\chi^2$ value against which all subsequent models were tested. Second, a model was run that constrained all factor loadings to be equivalent across both groups ($\chi^2 = 442.0, \text{df} = 40$). Results of the $\chi^2$ difference test suggest that some of the equality constraints did not hold across the two groups ($\chi^2\text{ diff} = 400.2, \text{df} = 6, p < .001$). Third, individual factor loadings were tested for noninvariance. As parameters were found to be equivalent across groups, their equality constraints were retained for further tests. Results of $\chi^2$ difference tests indicate that the only factor loading that was invariant across the groups was the Associations subtest ($\chi^2\text{ diff} = 4.3, \text{df} = 2, p$
> .10). These results should obviously be interpreted with caution, as the baseline models across the two groups contained different overall factor structures and violated certain assumptions of invariance testing (Werts, Rock, Linn, & Joreskog, 1976).

**Evaluation of Structural Invariance**

The structural regression coefficients of the SEM models in both samples were evaluated for invariance. Since the path between Fluency and Reading Comprehension was not significantly different from zero in the seventh-grade sample, the path between Breadth/Depth and Reading Comprehension in the seventh-grade sample was compared to the path between General Word Knowledge and Reading Comprehension in the third-grade sample. A baseline model was run in which all parameters were freely estimated for both groups ($\chi^2 = 41.8$, df = 34). The $\chi^2$ value of this model was compared to the $\chi^2$ value of a model in which the structural regression coefficients were constrained to be equal ($\chi^2 = 46.0$, df = 35). Results indicate that the paths are not invariant across the two samples ($\chi^2$ diff = 4.2, df = 1, $p < .05$). The dimension of word knowledge that was related to reading comprehension accounted for less variance in reading comprehension in the seventh-grade sample.

**Evaluation of Latent Means**

Testing of latent mean structures allows for examination of group differences in the means of the latent constructs (Bryne, 2001). The use of two different age samples in the present study implies that there should be growth in the means of the underlying constructs from third- to seventh-grade. Since different factor structures emerged in the two samples, testing for the invariance of mean structures was not possible for the factors of word knowledge. Although the factor of Reading Comprehension was present in both the third- and seventh-grade models, evaluation of differences in latent means was not possible because standardized variables were used for the measurement of this latent construct.
<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. EOWPVT</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. ROWPVT</td>
<td>.73</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Associations</td>
<td>.63</td>
<td>.56</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Multiple Contexts</td>
<td>.70</td>
<td>.62</td>
<td>.55</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Retrieval Fluency</td>
<td>.32</td>
<td>.32</td>
<td>.28</td>
<td>.30</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test of Word Finding</td>
<td>.71</td>
<td>.54</td>
<td>.47</td>
<td>.55</td>
<td>.22</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Decision Speed</td>
<td>.20</td>
<td>.18</td>
<td>.21</td>
<td>.16</td>
<td>.38</td>
<td>.14**</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Visual Matching</td>
<td>.11*</td>
<td>.09*</td>
<td>.22</td>
<td>.06*</td>
<td>.37</td>
<td>.09*</td>
<td>.54</td>
<td>___</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. FCAT-SSS Reading</td>
<td>.67</td>
<td>.61</td>
<td>.58</td>
<td>.67</td>
<td>.31</td>
<td>.52</td>
<td>.19</td>
<td>.22</td>
<td>___</td>
<td></td>
</tr>
<tr>
<td>10. SAT-10 Reading</td>
<td>.66</td>
<td>.57</td>
<td>.61</td>
<td>.65</td>
<td>.30</td>
<td>.50</td>
<td>.21</td>
<td>.31</td>
<td>.79</td>
<td>___</td>
</tr>
</tbody>
</table>

| Raw Score Mean                        | 92.23| 96.98| 12.92| 10.22| 50.24| 6.35 | 26.92| 36.77| ___  | ___  |
| Raw Score Standard Deviation          | 15.52| 12.71| 2.08 | 4.93 | 10.50| 3.62 | 4.54 | 5.16 | ___  | ___  |
| Maximum Possible                      | 170  | 170  | 15   | 32   | ___  | 22   | 40   | 60   | 500  | 834  |
| Standard Score Mean                   | 101.97| 101.16| 99.45| 10.74| 97.52| ___  | 105.25| 105.25| 332.03| 651.82|
| Standard Score Standard Deviation     | 15.78| 12.83| 14.96| 3.13 | 12.01| ___  | 14.72| 15.75| 66.84| 49.09|
| Reliability                           | .95a | .94a | .72a | .86a | .68b | .74a | .90b | .96b | ___  | ___  |

Note. N = 264.
* p > .10, ** p < .05. All other correlations significant at p < .01.
a Internal consistency reliability  b Reliability calculated using Rasch analysis
Table 4
Descriptive Statistics and Intercorrelations for All Observed Variables-7th Grade

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. EOWPVT</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. ROWPVT</td>
<td>.74</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Associations</td>
<td>.63</td>
<td>.56</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Multiple Contexts</td>
<td>.74</td>
<td>.68</td>
<td>.58</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Retrieval Fluency</td>
<td>.18*</td>
<td>.14*</td>
<td>.16*</td>
<td>.18**</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test of Adolescent Word Finding</td>
<td>.57</td>
<td>.34</td>
<td>.34</td>
<td>.42</td>
<td>.32</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Decision Speed</td>
<td>.14*</td>
<td>.10*</td>
<td>.08*</td>
<td>.05*</td>
<td>.32</td>
<td>.24</td>
<td>___</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Visual Matching</td>
<td>.15**</td>
<td>.11*</td>
<td>.16**</td>
<td>.19**</td>
<td>.26</td>
<td>.14*</td>
<td>.46</td>
<td>___</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. FCAT-SSS Reading</td>
<td>.62</td>
<td>.52</td>
<td>.59</td>
<td>.65</td>
<td>.25</td>
<td>.33</td>
<td>.26</td>
<td>.32</td>
<td>___</td>
<td></td>
</tr>
<tr>
<td>10. SAT-10 Reading</td>
<td>.62</td>
<td>.50</td>
<td>.56</td>
<td>.61</td>
<td>.29</td>
<td>.35</td>
<td>.25</td>
<td>.38</td>
<td>.80</td>
<td>___</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Raw Score</th>
<th>Mean</th>
<th>Raw Score</th>
<th>Standard Deviation</th>
<th>Maximum Possible</th>
<th>Standard Score</th>
<th>Mean</th>
<th>Standard Score</th>
<th>Standard Deviation</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>114.53</td>
<td>118.02</td>
<td>11.69</td>
<td>16.99</td>
<td>65.55</td>
<td>16.40</td>
<td>33.98</td>
<td>47.72</td>
<td>___</td>
</tr>
<tr>
<td></td>
<td>13.96</td>
<td>15.65</td>
<td>2.24</td>
<td>6.46</td>
<td>13.12</td>
<td>6.31</td>
<td>4.01</td>
<td>5.29</td>
<td>___</td>
</tr>
<tr>
<td></td>
<td>170</td>
<td>170</td>
<td>15</td>
<td>32</td>
<td>___</td>
<td>37</td>
<td>40</td>
<td>60</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>98.75</td>
<td>97.33</td>
<td>97.13</td>
<td>8.45</td>
<td>96.71</td>
<td>___</td>
<td>101.49</td>
<td>99.71</td>
<td>330.30</td>
</tr>
<tr>
<td></td>
<td>12.83</td>
<td>13.45</td>
<td>12.88</td>
<td>3.09</td>
<td>12.54</td>
<td>___</td>
<td>12.69</td>
<td>14.44</td>
<td>50.41</td>
</tr>
</tbody>
</table>

*Note.* N = 165.

*p > .10,* **p < .05. All other correlations significant at *p < .01.

*Internal consistency reliability b Reliability calculated using Rasch analysis*
<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$</th>
<th>$\chi^2$/df</th>
<th>RMSEA</th>
<th>AGFI</th>
<th>TLI</th>
<th>CFI</th>
<th>$\chi^2$ difference $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Three-Factor Model: Breadth, Depth, Fluency $^a$</td>
<td>9.98</td>
<td>6</td>
<td>.13</td>
<td>1.66</td>
<td>.05</td>
<td>.96</td>
<td>.99</td>
<td>.99</td>
<td></td>
</tr>
<tr>
<td>2. Two-Factor Model: Fluency, Breadth/Depth $^a$</td>
<td>11.93</td>
<td>8</td>
<td>.15</td>
<td>1.49</td>
<td>.04</td>
<td>.96</td>
<td>.99</td>
<td>.99</td>
<td></td>
</tr>
<tr>
<td>3. Two-Factor Model: Breadth, Depth/Fluency $^a$</td>
<td>13.38</td>
<td>8</td>
<td>.10</td>
<td>1.67</td>
<td>.05</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
<td></td>
</tr>
<tr>
<td>4. Two-Factor Model: Depth, Breadth/Fluency</td>
<td>11.89</td>
<td>8</td>
<td>.16</td>
<td>1.49</td>
<td>.04</td>
<td>.96</td>
<td>.99</td>
<td>.99</td>
<td></td>
</tr>
<tr>
<td>5. One-Factor Model: General Word Knowledge</td>
<td>13.50</td>
<td>9</td>
<td>.14</td>
<td>1.50</td>
<td>.04</td>
<td>.96</td>
<td>.99</td>
<td>.99</td>
<td>(4 vs.5), 1.62$^*$</td>
</tr>
<tr>
<td>6. One-Factor Model with Correlated Errors</td>
<td>2.79</td>
<td>8</td>
<td>.95</td>
<td>.35</td>
<td>.00</td>
<td>.99</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>7. One-Factor Structural Model</td>
<td>18.95</td>
<td>18</td>
<td>.40</td>
<td>1.05</td>
<td>.01</td>
<td>___</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Solution not admissible

$p > .10$
### Table 6
*Model Fit Indices - 7th Grade*

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$</th>
<th>$\chi^2$/df</th>
<th>RMSEA</th>
<th>AGFI</th>
<th>TLI</th>
<th>CFI</th>
<th>$\chi^2$ difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Three-Factor Model: Breadth, Depth, Fluency</td>
<td>17.21</td>
<td>6</td>
<td>.01</td>
<td>2.87</td>
<td>.11</td>
<td>.89</td>
<td>.94</td>
<td>.98</td>
<td></td>
</tr>
<tr>
<td>2. Two-Factor Model: Fluency, Breadth/Depth</td>
<td>20.06</td>
<td>8</td>
<td>.01</td>
<td>2.51</td>
<td>.10</td>
<td>.90</td>
<td>.95</td>
<td>.97</td>
<td>(2 vs. 1), 2.86*</td>
</tr>
<tr>
<td>3. Two-Factor Model: Breadth, Depth/Fluency</td>
<td>30.42</td>
<td>8</td>
<td>.00</td>
<td>3.80</td>
<td>.13</td>
<td>.85</td>
<td>.91</td>
<td>.95</td>
<td>(3 vs. 1), 13.21**</td>
</tr>
<tr>
<td>4. Two-Factor Model: Depth, Breadth/Fluency</td>
<td>29.17</td>
<td>8</td>
<td>.00</td>
<td>3.64</td>
<td>.13</td>
<td>.86</td>
<td>.91</td>
<td>.95</td>
<td>(4 vs. 1), 11.97**</td>
</tr>
<tr>
<td>5. One-Factor Model: General Word Knowledge</td>
<td>30.44</td>
<td>9</td>
<td>.00</td>
<td>3.38</td>
<td>.12</td>
<td>.87</td>
<td>.92</td>
<td>.95</td>
<td>(5 vs. 1), 13.24**</td>
</tr>
<tr>
<td>6. Fluency, Breadth/Depth with Correlated Error</td>
<td>3.84</td>
<td>7</td>
<td>.80</td>
<td>0.548</td>
<td>.00</td>
<td>.98</td>
<td>1.01</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>7. Two-Factor Structural Model: Fluency, Breadth/Depth</td>
<td>22.86</td>
<td>16</td>
<td>.12</td>
<td>1.43</td>
<td>.05</td>
<td>___</td>
<td>.98</td>
<td>.99</td>
<td></td>
</tr>
</tbody>
</table>

*p > .10, **p < .01
Figure 1. Three-factor measurement model of Breadth, Depth, and Fluency in the third-grade sample. All regression weights and covariances are significant ($p < .001$). eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; twf = Test of Word Finding.
Figure 2. Two-factor measurement model of Fluency and Breadth/Depth in the third-grade sample. All regression weights and covariances are significant ($p < .001$). eowpt = Expressive One-Word Picture Vocabulary Test; rowpt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; twf = Test of Word Finding.
Figure 3. Two-factor measurement model of Breadth and Depth/Fluency in the third-grade sample. All regression weights and covariances are significant ($p < .001$). eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; twf = Test of Word Finding.
Figure 4. Two-factor measurement model of Depth and Breadth/Fluency in the third-grade sample. All regression weights and covariances are significant ($p < .001$). eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; twf = Test of Word Finding.
Figure 5. One-factor measurement model of General Word Knowledge in the third-grade sample. All regression weights are significant ($p < .001$). eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; twf = Test of Word Finding.
Figure 6. Three-factor measurement model of Breadth, Depth, and Fluency in the seventh-grade sample. All regression weights and covariances are significant ($p < .001$).
eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; tawf = Test of Adolescent/Adult Word Finding.
Figure 7. Two-factor measurement model of Fluency and Breadth/Depth in the seventh-grade sample. All regression weights and covariances are significant ($p < .001$). eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; tawf = Test of Adolescent/Adult Word Finding.
Figure 8. Two-factor measurement model of Breadth and Depth/Fluency in the seventh-grade sample. All regression weights and covariances are significant ($p < .001$). eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; tawf = Test of Adolescent/Adult Word Finding.
Figure 9. Two-factor measurement model of Depth and Breath/Fluency in the seventh-grade sample. All regression weights and covariances are significant ($p < .001$). eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; tawf = Test of Adolescent/Adult Word Finding.
Figure 10. One-factor measurement model of General Word Knowledge in the seventh-grade sample. All regression weights are significant ($p < .001$). eowpv = Expressive One-Word Picture Vocabulary Test; rowpv = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; tawf = Test of Adolescent/Adult Word Finding.
Figure 11. One-factor structural model in the third-grade sample. All regression weights and covariances are significant ($p < .001$). eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; associations = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; twf = Test of Word Finding; sss = Florida Comprehensive Assessment Test- Sunshine State Standards; nrt = Stanford Achievement Test- Tenth Edition.
Figure 12. One-factor structural model with Processing Speed in the third-grade sample. * $p < .001$  ** $p < .05$. 
Figure 13. One-factor structural model with FCAT-SSS in the third-grade sample. The regression weight is significant ($p < .001$).
Figure 14. One-factor structural model with SAT-10 in the third-grade sample. The regression weight is significant ($p < .001$).
Figure 15. Two-factor structural model in the seventh-grade sample. * \( p > .10 \). All other regression weights and covariances significant \((p < .001)\). eowpvt = Expressive One-Word Picture Vocabulary Test; rowpvt = Receptive One-Word Picture Vocabulary Test; assoc = Associations subtest of The WORD Test-2; multcont = Multiple Contexts subtest of the Test of Word Knowledge; retfluency = Retrieval Fluency subtest of the Woodcock Johnson III Tests of Cognitive Abilities; tawf = Test of Adolescent/Adult Word Finding; sss = Florida Comprehensive Assessment Test- Sunshine State Standards; nrt = Stanford Achievement Test- Tenth Edition.
Figure 16. Two-factor structural model with Processing Speed in the seventh-grade sample. * $p < .001$  ** $p < .05$  *** $p > .10$. 
Figure 17. Structural model with Vocabulary Fluency as the only predictor variable in the seventh-grade sample.
*p < .001.
Figure 18. Structural model with Vocabulary Fluency and Processing Speed as predictor variables in the seventh-grade sample.

* $p < .001$  ** $p < .05$  *** $p > .10$.  

Fluency $\rightarrow$ Reading Comp $\rightarrow$ res

Fluency $\rightarrow$ Proc Speed

Proc Speed $\rightarrow$ Reading Comp

Fluency $\rightarrow$ .46**

Proc Speed $\rightarrow$ .22***

Reading Comp $\rightarrow$ .38

res
Figure 19. Two-factor structural model with FCAT-SSS in the seventh-grade sample.

* $p < .001$  *** $p > .10$.  

Fluency

Breadth/Depth

Fluency

Breadth/Depth

FCAT-SSS

e
Figure 20. Two-factor structural model with SAT-10 in the seventh-grade sample. $p < .001$ $*** p > .10$. 
DISCUSSION

One of the major goals of the current study was to advance the investigation of the structure of word knowledge by determining whether conceptualizations identifying three factors of breadth, depth, and fluency can be empirically supported. Findings from an earlier study (Tannenbaum et al, 2006) suggested that the three dimensions of word knowledge were not completely distinguishable from one another in a sample of third-grade children. Specifically, the authors of the earlier study found that a two-factor model of Breadth and Depth/Fluency provided the best fit to the data. The authors also found that breadth had a stronger relationship to reading comprehension than did depth/fluency; however, the two dimensions of word knowledge were highly correlated with one another (Tannenbaum et al, 2006). In the current study, another attempt was made to examine the robustness of a three-factor model of word knowledge in a demographically similar sample of third-grade children. The results indicated that a one-factor model of General Word Knowledge provided the best fit to the data. Taken together, the immediate theoretical implication of the results of the two studies is that the three factors of word knowledge are not empirically distinguishable from one another in third-grade children. The findings from the two studies also suggest that depth and fluency of word knowledge are not uniquely related to reading comprehension, beyond the variance they share with breadth of word knowledge. Therefore, the immediate practical implication from the findings of the two studies is that when considering the relationship between vocabulary and comprehension for instructional purposes in the schools, measures of breadth may be all that is currently necessary.

This is the first study to also examine the structure of word knowledge in a sample of seventh-grade children. Results indicated that a two-factor model of Fluency and Breadth/Depth provided the best fit to the data. Although these findings provide additional support for the lack of independence of the three factors of word knowledge, it is difficult to draw firm conclusions because these findings have not yet been replicated in another sample of seventh-grade children. Moreover, problems with the measurement of vocabulary fluency limit the inferences that can be made about the measurement of
word knowledge in seventh-grade children. These measurement problems are a limitation of the current study and will be addressed in further detail later.

As previously mentioned, one of the goals of this study was to replicate the findings of the Tannenbaum et al study (2006) in which a two-factor model of Breadth and Depth/Fluency emerged as the best-fitting model of vocabulary knowledge in a sample of third-grade children. Although the current study used a demographically similar sample of third-grade children, a one-factor model of General Word Knowledge provided the best fit to the sample data. Moreover, the two-factor model of Breadth and Depth/Fluency was not even admissible in the third-grade sample in this study because the factors of word knowledge were so highly correlated. Taken together, the results of these two studies provide initial evidence that the three factors of word knowledge are not robust in third-grade children, and the structure of word knowledge depends critically on the measures employed in the study, or on relatively subtle and unknown differences in the characteristics of the students in the two random samples of third-grade students. In other words, the dimensions of word knowledge may be so highly related to one another that the different results obtained in these two studies are due to subtle measurement artifacts. Evidence for this argument comes from the large correlations found between the different factors of word knowledge in both studies. In the current study, the correlations between the latent variables ranged from .96 to 1.12, consistent with the finding that the dimensions of word knowledge were not completely distinguishable from one another. Although Breadth and Depth/Fluency were distinguishable from one another in the earlier study, they were very highly correlated ($r = .86$) (Tannenbaum et al, 2006).

As previously noted, a second goal of this study was to examine the structure of word knowledge in a sample of seventh-grade children. Results of confirmatory factor analyses showed that a two-factor model of Fluency and Breadth/Depth best represented the dimensions of word knowledge in a seventh-grade sample. Although the two dimensions of word knowledge were correlated with one another ($r = .59$), they were not as strongly related as the dimensions of word knowledge in the Tannenbaum et al study (2006) of third-grade students. One possible explanation for this difference is that the correlation in the current study was attenuated because vocabulary fluency was not adequately measured. In particular, the correlations between the two measures of
vocabulary fluency in the third-grade ($r = .22$) and seventh-grade ($r = .32$) samples were lower than the correlations of the companion measures of vocabulary breadth and depth, indicating that the measurement of vocabulary fluency was less coherent than the measurement of the other dimensions of word knowledge. Moreover, the weak correlations between the Retrieval Fluency task and the other measures of word knowledge suggest that this measure in particular did not function well as a measure of vocabulary knowledge. In the third-grade sample, the average correlation between Retrieval Fluency and the other measures of word knowledge ($r = .30$) was significantly weaker ($p<.05$) than the average correlation between the TWF and the other measures of word knowledge ($r = .57$). Although the average correlation between Retrieval Fluency and the other measures of word knowledge ($r = .16$) was smaller than the average correlation between the TAWF and the other measures of word knowledge ($r = .42$) in the seventh-grade sample, the difference was not statistically significant. Examination of the factor loading of the Retrieval Fluency task provides further evidence that this task did not perform well as a measure of vocabulary knowledge in both samples. Specifically, the factor loadings of the Retrieval Fluency task in the third- and seventh-grade samples were .36 and .34, respectively. The factor loadings of the other measures of word knowledge ranged from .69-.93 in the third-grade sample and .69-.94 in the seventh-grade sample.

The overall structure of word knowledge was found to be different across the third- and seventh-grade samples. One possible explanation for the different results is that the factors of word knowledge become more differentiated with age. This finding is consistent with age-differentiation hypotheses that propose that ability structures are integrated in early childhood and undergo a process of differentiation during late childhood and adolescence (Baltes, Cornelius, Spiro, Nesselroade, & Willis, 1980; Bos, 1996; Burt, 1954; Garrett, 1946; Juan-Espinosa, Garcia, Colom, & Abad, 2000; Juan-Espinosa, Garcia, Escorial, Rebollo, Colom, & Abad, 2002).

Another possible explanation for the finding of different overall factor structures in the two grade-level samples is that the measures of word knowledge used in the current study functioned differently at third- and seventh-grades. Specifically, the finding of different overall factor structures across multiple samples can serve as an indication
that the tasks are not measuring the same constructs (Labouvie, 1980). Additional evidence for the argument that the measures of word knowledge were operating differently in the two samples comes from the tests of measurement equivalence, where the majority of the measures of word knowledge were not found to be invariant across the two samples. Unequal factor loadings suggest that the tasks are not identically related to the constructs across the two samples (Labouvie, 1980). The finding of variant factor loadings also signals that caution should be used in assuming that the factors of word knowledge become more differentiated with age since the factors are not fully comparable across the two samples (Babcock, Laguna, & Roesch, 1997).

A third goal of this study was to evaluate the relationships between the dimensions of word knowledge and reading comprehension in the two samples. All of the third-grade analyses demonstrate that general word knowledge is very highly related to reading comprehension. This finding is consistent with the results of the Tannenbaum et al study (2006). Although more variance in reading comprehension was accounted for by the dimension of General Word Knowledge ($R^2 = .76$) than the combined contribution of the dimensions of Breadth and Depth/Fluency in the previous study ($R^2 = .62$), this difference is not statistically reliable. Even though the difference is not statistically significant, one aspect of the data that is consistent with the finding of higher correlations in the current study is that the variability in performance on the reading comprehension measures was higher than in the previous study. In the most recent study, the standard deviations of FCAT-SSS and SAT-10 scores were 66.8 and 49.1, respectively. In the earlier study, these same standard deviations were 53.2 and 38.7.

Large correlations were also found between the dimensions of word knowledge and reading comprehension in the seventh-grade sample. The correlation between Breadth/Depth and Reading Comprehension ($r = .80$) was substantially higher than the correlation between Fluency and Reading Comprehension ($r = .52$). Likewise, Breadth/Depth was the only dimension of word knowledge in the seventh-grade sample that accounted for unique variance in Reading Comprehension (28%). These results demonstrate that Breadth/Depth had a stronger relationship to reading comprehension than did Fluency in the seventh-grade sample. Some researchers have suggested that the vocabulary load present in many middle school texts inherently draws for knowledge of a
broad range of words and thorough understanding of known words (Alvermann & Phelps, 2002; Beck et al, 2002; Harmon, Hedrick, & Wood, 2005). The results of the current study provide initial evidence that breadth and depth of word knowledge are particularly important in the comprehension of texts in the seventh-grade.

Although the measures of the dimension of vocabulary fluency provided less coherent measurement of the underlying construct that did the measures of the other two vocabulary constructs, it is clear from other analyses that the measures of vocabulary fluency were measuring something beyond mere processing speed in the seventh-grade sample. The structural coefficient for Vocabulary Fluency was .60 when it was the only predictor variable included in the structural model. This value decreased to .46 when Processing Speed was added to the model; however, this reduction was not statistically reliable. Moreover, the path still remained significantly different from zero. In contrast, the structural coefficient for Vocabulary Fluency was not significantly different from zero when the vocabulary dimension of Breadth/Depth was included in the model. The reduction in the structural coefficient for Vocabulary Fluency was statistically reliable with the addition of Breadth/Depth. Taken together, these results indicate that vocabulary fluency’s overlap with breadth/depth is the most important reason that it is related to reading comprehension. This suggests that the vocabulary fluency measures were, indeed, measuring something about the extent of students’ vocabulary—what the current study failed to show is that there is something about fluency of access to word meanings that is uniquely related to reading comprehension beyond breadth and depth of word knowledge.

A final goal of the current study was to examine equivalence of measurement components, structural components, and latent means across the two samples. The results from these analyses were limited because different overall factor structures emerged in the third- and seventh-grade samples and measurement equivalence was not found for the majority of the measures of word knowledge. Measurement equivalence was found for the FCAT-SSS and SAT-10, suggesting that these measures were related identically to the construct of reading comprehension across both samples. Testing of latent means was not possible for the dimensions of word knowledge because different overall factor structures emerged in the two samples. Moreover, standard scores were used for the
measurement of reading comprehension, making it impossible to examine growth in the mean of the underlying construct. Results of the invariance testing of structural parameters corroborated the finding that the dimensions of word knowledge accounted for less variance in reading comprehension in the seventh-grade sample than the third-grade sample.

The findings from the current study highlight the need for development of better measures of vocabulary depth and fluency, as they are conceptualized in the current framework. As previously mentioned, the factor loadings of the Retrieval Fluency task suggest that it functioned poorly as a measure of vocabulary fluency in both samples. Moreover, the reliabilities of the two measures of vocabulary fluency and the Associations task were relatively low in both samples. Ninety-five percent confidence intervals were created for average reliability estimates of the measures of breadth, depth, and fluency of word knowledge. The confidence intervals were then examined for overlap to determine if the reliability estimates were statistically different from one another. This procedure revealed that the reliability of the measures of breadth in the third-grade ($\alpha = .945$) and seventh-grade ($\alpha = .945$) samples was statistically stronger than the depth and fluency measures. The reliability of the measures of depth in the third-grade ($\alpha = .785$) and seventh-grade ($\alpha = .755$) samples were not statistically different than the reliability of the measures of fluency in the third-grade ($\alpha = .71$) and seventh-grade ($\alpha = .825$) samples. In the Tannenbaum et al study (2006), the authors also found that the reliabilities of the measures of vocabulary depth and fluency were statistically weaker than the reliabilities of the measures of vocabulary breadth. Although measurement error is taken into account in analyses with latent variables, estimates are more precise when psychometrically sound measures are used (Kline, 2005).

Measurement of depth and fluency of word knowledge, as they are conceptualized in the current framework, is also complicated by the lack of measures that exist. In the current study, the standard administration procedures of one of the measures of vocabulary fluency had to be modified to better represent the underlying construct because no other tests were found that adequately measured vocabulary fluency.

A possible approach to the measurement of these constructs in the future may be to develop a test that measures breadth, depth, and fluency of word knowledge using the
same words. This type of approach was used in a recent unpublished master’s thesis (Phythian-Sence, 2006), where the author used one set of words to measure definitional word knowledge, contextual word knowledge, morphological word knowledge, and conceptual word knowledge. Although many researchers have discussed different conceptualizations of word knowledge, few studies have been conducted using identical word sets to measure different types of word knowledge.

The use of tests with non-identical word sets is another limitation of the current study. Using different tasks to measure the three dimensions of word knowledge requires the assumption that the target words on the tests were selected in comparable manners. Strategies used for item selection were described by all of the test developers in the manuals; however, the quality of information provided about item selection procedures varied. Several common sources were used to select items for some of the tests used in this study. For example, some of the test developers used The Living Word Vocabulary (Dale & O’Rourke, 1981), Basic Reading Vocabularies (Harris & Jacobson, 1982), and the EDL Core Vocabularies in Reading, Mathematics, Science, and Social Studies (Taylor, Frackenpohl, & White, 1989) to guide their item selection process. Other sources used include The Educator’s Word Frequency Guide (Zeno, 1995), The Reading Teacher’s Book of Lists (Fry et al, 1993), the Computational Analysis of Present-Day American English (Kucera & Francis, 1967), and the American Heritage Word Frequency Book (Carroll, Davies, & Richman, 1971). The most common sources used for item selection were word frequency guides, grade-level curriculum materials, and item difficulty levels. Although these strategies were used for the item selection process of several tests employed in this study, a few of the test developers provided limited information about the test development process. Thus, caution should be used in assuming that comparable words were used to assess vocabulary breadth, depth, and fluency in the current study.

A final limitation of this study is that the use of structural equation modeling does not necessarily signify that casual claims can be made about relationships between constructs. Moreover, the data for this study were not collected in a longitudinal manner, so it is inappropriate to make any causal assumptions about the relationships between word knowledge and reading comprehension. Likewise, for most structural models there
will also be other models that are indistinguishable from the original model in terms of fit to the sample covariance matrix (MacCallum, Wegener, Uchino, & Fabrigar, 1993). For example, the arrows between the dimensions of word knowledge and reading comprehension could be reversed, and the covariance between Fluency and Breadth/Depth in the seventh-grade sample could be changed to a direct path. It is important to consider equivalent models because other plausible explanations of the data may exist (MacCallum, Wegener, Uchino, & Fabrigar, 1993). Future research should examine plausible equivalent models and casual relations among the dimensions of word knowledge and reading comprehension.
APPENDIX A

FLORIDA STATE UNIVERSITY HUMAN SUBJECTS COMMITTEE
APPROVAL LETTER
Office of the Vice President For Research  
Human Subjects Committee  
Tallahassee, Florida 32306-2742  
(850) 644-8633  FAX (850) 644-4392

APPROVAL MEMORANDUM

Date: 11/2/2006

To:  
Kendra Tannenbaum  
FCRR, 227 N. Bronough Street, Ste. 7250  
Tallahassee, FL 32301

Dept.: PSYCHOLOGY DEPARTMENT  

From: Thomas L. Jacobson, Chair  

Re: Use of Human Subjects in Research  
Relationships Between Measures of Word Knowledge and Reading Comprehension in  
Third-Grade and Seventh-Grade Children

The forms that you submitted to this office in regard to the use of human subjects in the proposal referenced above have been reviewed by the Human Subjects Committee at its meeting on 10/11/2006. Your project was approved by the Committee.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals which may be required.

If the project has not been completed by 10/10/2007 you must request renewed approval for continuation of the project.

You are advised that any change in protocol in this project must be approved by resubmission of the project to the Committee for approval. The principal investigator must promptly report, in writing, any unexpected problems causing risks to research subjects or others.

By copy of this memorandum, the chairman of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department, and should review protocols of such investigations as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Protection from Research Risks. The Assurance Number is IRB00000446.

cc: Joseph Torgensen  
HSC No. 2006.0903
APPENDIX B

SAMPLE INFORMED CONSENT LETTER
Dear Parent,

We are sending this letter to you because we would like permission to include your child in a study we are conducting at your elementary or middle school. The purpose of the study is to examine the relationships between several measures of word knowledge and reading comprehension in third-grade and seventh-grade children. This information will help provide knowledge on how teachers can best teach vocabulary and what type of word knowledge is most highly related to success on the FCAT. We are asking **all** children in your child’s classroom to participate.

Your child’s participation will involve completing a series of individually administered measures. These measures will be given to **all** children in your child’s classroom whose parents or guardians consent. The administrator is a highly trained student at FSU. The administration should require two thirty-minute sessions with your child. These measures will assess your child’s word knowledge and general verbal ability. We will schedule the sessions at a time that is least disruptive to ongoing classroom work.

One way we will examine the relationships between word knowledge and reading comprehension is to compare your child’s performance on these measures with several routine tests given by your child’s school. Thus, we will need information about your child’s scores on the SAT-10 measure of reading comprehension and the FCAT, which will require us to access your child’s identification numbers.

Your participation, as well as that of your child, in this study is voluntary. If you or your child chooses not to participate or to withdraw from the study at any time, there will be no penalty and it will not affect your child’s grade. We will provide your child with a small prize after participating in the study. The results of the research study may be published, but your child’s name will never be used in any reports of the data. We will only report group data. Information obtained during the course of the study will remain confidential, to the extent allowed by law.

The project has the approval of Leon County Research Advisory Board as well as that of your school principal. If you have any questions concerning this study or your child’s participation, please call Kendra Tannenbaum (645-7135). If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the FSU Human Subjects Committee, Institutional Review Board, through the Vice President for the Office of Research (644-9694).

Sincerely,

Kendra R. Tannenbaum, M.S.  
Project Coordinator

Joseph K. Torgesen, Ph.D.  
Director of the Florida Center for Reading Research

PERMISSION FORM

I GIVE consent for my child (print child’s name here) ________________________________ to participate in the above study.

________________________
Signature of child’s parent or guardian

OR

I DO NOT want my child (print child’s name here) ________________________________ to participate in the above study.

________________________
Signature of child’s parent or guardian
Child Assent

Hello, ____________, my name is _______________ and I am a student at Florida State University. I would like your help in a project that I am doing. I’m asking all the kids in your classroom to participate whose parents provided permission. For my project, you are going to participate in several different activities, and we will be meeting with you two times. I’ll be working with you today, and one of my friends will be working with you in about a week. You won’t be receiving a grade on the activities. These activities are similar to things you do at school. For example, I will be asking you to point to certain pictures and tell me what certain words mean. Some of the questions are hard and you may not know the answer, but that’s okay. You just try your best, and if you don’t want to continue you let me know and we can stop at any time. So do you want to participate and help me with my project?

If yes- Great. Let’s get started.

If no- The child will be excused from the study and escorted back to the classroom.
APPENDIX D

LEON COUNTY RESEARCH ADVISORY BOARD APPROVAL LETTER
October 31, 2006

Ms. Kendra R. Tannenbaum
Florida Center for Reading Research
227 N. Bronough St. Suite 7250
Tallahassee, FL 32301

Dear Ms. Tannenbaum,

The Leon County Schools Research Review Board has determined that the findings of your proposed study could be pertinent to our efforts and so we are approving your request to conduct research. This includes your email clarifications of October 3 and 10, 2006. We also expect that you will coordinate the timing of data collection so that it does not interfere with FCAT. Thank you for sending the results of your last study by the same title.

Your research request is approved for the period of October 2006 through September 2007. Should you desire to extend your research efforts after this period of time, you must submit (a) a progress report, (b) preliminary results of your research, and (c) a request for renewed approval for continuation. Any significant changes or amendments to the procedures or design of this study must be approved by resubmitting the request for research to the Research Review Board.

Approval by the Research Review Board does not in itself constitute permission to carry out the research. You may now contact principals of the schools in your study. The principal has the final decision relative to research at each school. It is your responsibility to return the enclosed "Principal's Consent for Research Participation," signed by the principal(s) of the school(s) to be involved, prior to the start of any research. Receipt of this form by this office will complete the approval process.

Since your research study involves direct contact with students, the background check policy requires all researcher(s) to be fingerprinted for clearance. It is the responsibility of the applicant(s) to complete all required documentation prior to the beginning the study.

In order to obtain the requested FCAT reading scores, our office must receive the signed consent forms and an electronic file containing all the students' names and their Leon County Schools identification number.

Leon County Schools is approving your research partly for the potential benefit of information to the district; therefore, it is important that you send this office one copy of your results and discussion when your study is complete. We will place information from your study in our research library and annotated listing of conducted research. We look forward to receiving your results.

Please feel free to phone me (850.488.7007) if I may be of further assistance.

Sincerely,

Margarita S. Southard, Ph.D.
Chairperson, Research Review Board

Ms. Kendra R. Tannenbaum,
Florida Center for Reading Research
227 N. Bronough St. Suite 7250
Tallahassee, FL 32301
REFERENCES


BIOGRAPHICAL SKETCH
Kendra R. Tannenbaum

EDUCATION

Dates: September 2005- August 2008; Doctor of Philosophy
     Florida State University, Tallahassee, Florida
     Major Area: Clinical Psychology
     Major Professor: Joseph K. Torgesen
     Dissertation: Relationships between measures of word knowledge and reading comprehension in third- and seventh-grade children.

Date: September 2005; Passed Preliminary Doctoral Exam

Dates: September 2002-April 2005; Master of Science
     Florida State University, Tallahassee, Florida
     Master’s Thesis: Relationships between measures of word knowledge and reading comprehension in third grade children

Dates: September 1997-May 2001; Bachelor of Arts
     Johns Hopkins University, Baltimore, Maryland
     Major: Behavioral Biology

RESEARCH EXPERIENCES

Graduate Research Assistant, Florida Center for Reading Research, Florida State University, Department of Psychology, Tallahassee, Florida.
     Duties: Expanded knowledge of vocabulary development, assessment, and instruction; Conducted educational assessments of children; Item development for diagnostic tests; Conducted, organized, and oversaw data collection; Conducted data entry and analysis using Microsoft Excel, SPSS, and AMOS; Coordinated research projects; Supervised undergraduate students.
     Supervisor: Joseph Torgesen, Ph.D.

Research Data Assistant, Johns Hopkins Medical Institutions, Department of Child and Adolescent Psychiatry, Baltimore, Maryland.
     Duties: Recruited participants from local clinics and schools; Tested normally developing children and children with Autism Spectrum Disorders in their autobiographical memory and suggestibility; Data entry and analysis; Organizational aspects of laboratory; Oversaw and coordinated research project.
     Supervisor: Maggie Bruck, Ph.D.
Lab Assistant, Johns Hopkins University, Department of Cognitive Science, Baltimore, Maryland.  
Duties: Tested adult participants in their cognitive abilities to determine the effects of stroke; Data entry and analysis using Microsoft Excel; Compilation of spelling tests.  
Supervisor: Brenda Rapp, Ph.D.

PUBLICATIONS


EDITORIAL WORK


PRESENTATIONS


TEACHING EXPERIENCES

Directed Individual Study Supervisor, Florida State University, Department of Psychology. Tallahassee, Florida.
Duties: Supervised teams of 4-7 undergraduate students who were assisting with ongoing research projects; Led weekly meetings in which the students discussed selected articles relevant to vocabulary and reading comprehension; Trained students on reading assessment administration and conducting research in school systems; Provided tutorials on data management and statistics (e.g. ANOVA, regression, CFA, SEM) to enhance students’ research design skills and statistical knowledge.
Supervisor: Joseph Torgesen, Ph.D.

Teaching Assistant, Johns Hopkins University, Department of Sociology, Baltimore, Maryland.
Duties: Graded papers of undergraduate students; Led field trips to a juvenile detention center, Led a section of the class.

CLINICAL EXPERIENCES

Predoctoral Psychology Intern, Children’s Hospitals and Clinics of Minnesota, St Paul, Minnesota.
Dates: September 2007-August 2008
Duties: Provided outpatient therapy to children and adolescents with a variety of presenting problems through the use of empirically supported treatments; Conducted psychological evaluations; Conducted interdisciplinary team assessments; Provided consultations to the inpatient medical units; Provided consultations for the hematology/oncology service; Attended weekly meetings for the hematology/oncology service; Served as co-facilitator of treatment group for adolescent boys with ADHD and Disruptive Behavior Disorders; Spent one-half day per week at a pediatric primary care practice providing consultation to parents and medical staff; Attended weekly didactic seminars in addition to professional development and ethics seminars; Attended workshops on multicultural diversity and supervision; Attended medical grand rounds; Conducted four formal presentations; Attended four hours of weekly supervision; Completed appropriate paperwork (diagnostic assessments, treatment plans, progress notes).
Supervisors: Sarah Jerstad, Ph.D., LP, Denny Marvinney, Ph.D., LP, Michael Troy, Ph.D., LP, Jason Walker, Ph.D., LP
Director of Training: Sharon Berry, Ph.D., LP

Master’s Level Therapist, Child Crisis Stabilization Unit (CCSU), Apalachee Center, Tallahassee, Florida.
Duties: Provided crisis intervention services to children and adolescents in a psychiatric inpatient setting; Attended weekly supervision; Consulted and collaborated with case managers, psychiatrists, and nurses; Completed appropriate paperwork (i.e. progress notes, service logs for billing purposes).
Supervisors: Jay Reeve, Ph.D., Brent Lovett, M.D.

Master’s Level Therapist, S.A.F.E. (Supportive Alliance for Family Enrichment) Program, Apalachee Center, Tallahassee, Florida.
Duties: Provided in-home family therapy services to families whose children were at risk for placement into a more intensive level of care; Provided individual therapy to child and adolescent clients; Attended medication management meetings with clients; Attended weekly supervision and SAFE team meetings; Consulted and collaborated with case managers and psychiatrists; Developed and reviewed client service plans; Completed appropriate paperwork (i.e. progress notes, client service plans, service logs for billing purposes).
Supervisor: Jay Reeve, Ph.D.
Crisis Interventionist, Florida State University Crisis Management Unit, Tallahassee, Florida. Dates: August 2005-August 2007. Duties: Provided emergency services for mental health crises; Conducted crisis intervention and evaluation of psychotic, suicidal, and homicidal adults; Provided stabilization, referrals, and Baker Acts for emergency evaluations (averaged five 24-hour on-call shifts per month). Supervisor: Joyce Carbonell, Ph.D.

Psychological Trainee, Autism and Related Disorders Clinic, FSU Multidisciplinary Center, Tallahassee, Florida. Dates: May 2006-August 2006, August 2005-December 2005. Duties: Evaluated children and adolescents to determine the extent to which they displayed behaviors characteristic of Autism or an Autism Spectrum Disorder; Administered, scored, and interpreted assessment instruments; Completed psychological reports; Provided diagnostic feedback and referrals to parents and guardians; Attended weekly supervision; Completed appropriate paperwork. Supervisors: Anne Selvey, Ph.D., Lauren Hutto, Ph.D., Eve Wettstein, Ed.S., Director: Beverly Atkeson, Ph.D.

Psychological Trainee, ADHD Clinic, FSU Multidisciplinary Center, Tallahassee, Florida. Dates: January 2006-May 2006. Duties: Evaluated children and adolescents to determine the extent to which they displayed behaviors characteristic of Attention-Deficit/Hyperactivity Disorder; Administered, scored, and interpreted assessment instruments; Observed client in school setting; Interviewed client’s teacher; Completed psychological reports; Provided diagnostic feedback and referrals to parents and guardians; Attended weekly supervision; Completed appropriate paperwork. Supervisor: Kristen Schmidt, Ph.D. Director: Beverly Atkeson, Ph.D.


Graduate Therapist, FSU Psychology Clinic, Tallahassee, Florida. Dates: May 2003-August 2005. Duties: Provided outpatient-based therapy to adults, adolescents, and children with a variety of presenting problems through the use of empirically supported treatments; Conducted psycho-educational and behavioral assessments of adults,
adolescents, and children; Led a cognitive-behavioral therapy group for children with anxiety problems (Spring, 2005); Led two social skills groups for elementary school children with behavior problems (Spring 2004, Fall 2004); Led a Parent Management Training group for parents of children with behavior problems (Fall, 2004); Advertised clinic services in local elementary schools; Attended weekly supervision; Attended weekly didactic training; Conducted screening interviews and intake interviews; Completed appropriate paper work (i.e. progress notes, intake reports, treatment plan summaries, termination reports.

Supervisors: Thomas Joiner, Ph.D. & Ellen Berler, Ph.D.

*Mental Health Volunteer*, Sheppard Pratt Health System- Adolescent Residential Treatment Center, Baltimore, Maryland.

**Dates**: February 2001-June 2002.

**Duties**: Participated in unit meetings; Assisted in making rounds; Helped to meet patients’ needs; Co-led activity groups.

**Supervisor**: Mary Edwards, M.A.