

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

Kendra R. Tannenbaum, Joseph K. Torgesen,
and Richard K. Wagner

*Department of Psychology and Florida Center for Reading Research
Florida State University*

The relationship between word knowledge and reading comprehension has been well documented in research; however, the nature of this relationship remains unclear. Researchers usually distinguish between 2 aspects of an individual's word knowledge: breadth and depth. In addition to these 2 factors, it may be important to also consider fluency in the study of word knowledge. Two hundred and three 3rd-grade students took part in a study that examined the relationships between 3 dimensions of word knowledge and reading comprehension. Confirmatory factor analyses, structural equation modeling, and hierarchical regression analyses show that a 2-factor model of breadth and depth/fluency provides the best fit to the data. Breadth has a stronger relationship to reading comprehension than does depth/fluency; however, the 2 dimensions of word knowledge have significant overlapping variance that contributes to the prediction of reading comprehension.

One of the most important skills learned by young students is the ability to understand written text, which is usually referred to as reading comprehension. Comprehension of the information in text, or of the author's meaning, is the ultimate reason for reading. To comprehend larger units of text such as paragraphs and stories, a child must understand the smaller word units (National Institute of Child Health and Human Development, 2000). The relationship between word knowledge and

reading comprehension has been well documented in research through various kinds of correlational analyses as well as through true experiments.

Early factor analytic studies established vocabulary knowledge as one of the major factors in reading comprehension (Davis, 1944; Spearritt, 1972). For example, Davis factor analyzed tests of nine skills underlying reading comprehension and found two major factors: word knowledge and reasoning. These two factors accounted for 89% of the variance in reading comprehension. Spearritt reanalyzed Davis's data and found four main factors: word knowledge, drawing inferences from the content, following the structure of the passage, and recognizing a writer's intent and tone. Word knowledge was the best differentiated of these four skills.

The strength of the relationship between word knowledge and reading comprehension has also been well established in simple correlational studies over the past 50 years. Most studies report correlations between word knowledge and reading comprehension that range from .3 to .8. The range of the correlations varies depending on several factors, such as test format, age of the participants, range of vocabulary knowledge in the research sample, and the dimension of word knowledge measured by the test. Reading vocabulary measures correlate higher with reading comprehension than do measures of oral vocabulary (Hafner, Weaver, & Powell, 1970; Stanovich & Cunningham, 1992) because measures of reading vocabulary tap word-reading ability in addition to knowledge of word meanings. The age of the participants also appears to affect the correlation between vocabulary and reading comprehension. Snow (2002) found that the strength of the relationship between a kindergarten vocabulary measure and reading comprehension increased substantially as the children advanced in grade level. The correlations for first graders, fourth graders, and seventh graders were .45, .62, and .69, respectively. Finally, another factor that might affect the correlation between vocabulary and reading comprehension is the dimension of vocabulary being assessed. Measures such as the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1991) and the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) assess breadth of word knowledge. These measures assess the range of words about which the individual has some knowledge, whereas other measures of vocabulary can be used to assess depth (Anderson & Freebody, 1981; Beck, McKeown, & Kucan, 2002) and fluency (Kaminski et al., 2004) of word knowledge.

The presence of a relationship between vocabulary and reading comprehension has also been supported in research that manipulated level of word knowledge experimentally. For example, altering the difficulty of the words in a text makes the text easier or harder to comprehend (Marks, Doctorow, & Wittrock, 1974; Raphael, Myers, Tirre, Fritz, & Freebody, 1981). In one study, 15% of the words in several elementary school texts were replaced with either high-frequency (easy) or low-frequency (hard) synonyms. Reading comprehension was significantly increased with the high-frequency group and reduced in the low-frequency group (Marks et al., 1974). A potential problem in studies such as this is that word fre-

quency likely affects both knowledge of a word's meaning and how accurately the word is read in text. In other words, the reduction in comprehension that resulted from replacement of high-frequency words with low-frequency synonyms could be due to the words being read less well because they are encountered less in print.

Additional experimental evidence for the relationship between vocabulary and reading comprehension is available from studies that taught word meanings to children to enhance comprehension of text. Some of these experiments have used methods to enhance word knowledge that have had an impact on both word knowledge and comprehension (Beck, McCaslin, & McKeown, 1980; Beck & McKeown, 1983), whereas others have produced impacts on word knowledge but only marginal effects on comprehension (Jenkins, Pany, & Schreck, 1978, Experiment 3; Pany & Jenkins, 1977; Tuinman & Brady, 1974).

Beck and her colleagues hypothesized that for vocabulary instruction to affect reading comprehension, the instruction must go beyond establishing a definition for a word; it must develop complex, in-depth knowledge about the words being taught. Good vocabulary instruction therefore provides contextual information in addition to definitional information about a word. A child needs to develop knowledge of the core concept of the word and how the word is used in different contexts to develop flexible knowledge about a word that contributes to reading comprehension. In addition, instruction should provide multiple, varied encounters with the words being taught. Multiple repetitions help improve the speed of accessing the word's meaning (Beck et al., 2002). Studies that provided this rich vocabulary instruction showed gains in both accuracy of word knowledge and comprehension of text containing the taught words (Beck & McKeown, 1983; McKeown, Beck, Omanson, & Perfetti, 1983).

Researchers usually distinguish between two aspects of an individual's vocabulary knowledge: breadth and depth. *Breadth* of vocabulary refers to the size of the mental lexicon and is the number of words that have some level of meaning to the individual. However, breadth does not specifically address how well each of these words is known. It is a dimension defined by how many words have meaning for the individual (Anderson & Freebody, 1981). *Depth* refers to the richness of knowledge that the individual possesses about the words that are known. Depth is not an all-or-none concept. At the most basic level, a word can be recognized but not well defined. As more depth is acquired, this word can be defined in greater detail. Finally, relations can be made between the word and other words, multiple meanings of a word can be learned, and the word can be used in different contexts. As depth of word knowledge increases, words can be used more flexibly, and their meaning can be readily appreciated and accessed within multiple contexts (Anderson & Freebody, 1981; Beck et al., 2002; Stahl, 1998). In one study, artificial words were taught to college students either in a single context or varied contexts. Students who received instruction in varied contexts were significantly better at understanding the word overall and understanding the word in novel contexts (Nitsch, 1978).

In addition to breadth and depth, it may be important to also consider fluency in the study of word knowledge. Breadth and depth of vocabulary are often discussed in the scientific literature, but fluency is rarely included as a dimension of word knowledge. *Fluency* is defined as the rate at which the individual accesses the meaning of a word. As word meanings become reinforced through frequent exposures, the time it takes to access the word meanings decreases (Wolf, Miller, & Donnelly, 2000). Beck and her colleagues (2002) suggested that fluency of access to word meanings may be one of the important factors explaining differences in outcomes from vocabulary training studies. Vocabulary training that produces an impact on the fluency with which word meanings are accessed has an impact on reading comprehension, but vocabulary instruction that does not produce sufficient fluency of access frequently does not generalize to reading comprehension (cf. Jenkins et al., 1978).

This study examines three dimensions of word knowledge in their relationships to reading comprehension. Confirmatory factor analysis (CFA), structural equation modeling, and hierarchical regression analysis were used to (a) evaluate whether the three dimensions were distinguishable factors of word knowledge, (b) examine the strength of the relationships between the three proposed dimensions of word knowledge, and (c) examine the relationships between each dimension of word knowledge and reading comprehension.

METHOD

Participants

Data for this study were collected from 203 third-grade students from a suburban school district in a moderate-size, southern U.S. city. Although relationships between the vocabulary constructs examined in this study and reading comprehension might be different at other age levels, third graders were used in this initial study because the end of third grade is widely regarded as the end point for the development of early reading skills (Chall, 1983). It is also the first point at which measures of reading comprehension are typically used to establish accountability for school success in teaching reading to all students in the United States. The sample contained 104 girls (51.2%) and 99 boys (48.8%) who were randomly selected from four elementary schools. Students were not excluded from the study based on English-language proficiency or handicapped status. The four elementary schools served a socioeconomically diverse student population, with 20.2% of the sample qualifying for free or reduced-price lunch. The ethnic composition of this sample was generally representative of the community: 68.0% White, 18.2% Black, 7.4%

Asian, 2% Hispanic, and 4.4% Other. The sample ranged in age from 99 to 138 months ($M = 111.14$ months, $SD = 5.60$).

Vocabulary Measures

Two measures from each of the three proposed dimensions of word knowledge (breadth, depth, and fluency) were administered to each participant. The PPVT-III (Dunn & Dunn, 1997) and the Vocabulary subtest of the WISC-III (Wechsler, 1991) were used to assess breadth of word knowledge. The Multiple Meanings and Attributes subtests of the Language Processing Test-Revised (LPT-R; Richard & Hanner, 1995) were used to measure depth of word knowledge. The Word Use Fluency (WUF) subtest of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) and an experimenter-developed Semantic Category Fluency test were used to assess fluency of word knowledge.

PPVT-III. The PPVT-III (Dunn & Dunn, 1997) is a measure of receptive vocabulary in which the participant is required to select a picture that best depicts meaning of a word spoken by the examiner. Stimulus items are presented in ascending difficulty, and testing is stopped after a child misses 8 or more items in a 12-item set. Internal consistency reliability of the PPVT-III in this sample was .94.

WISC-III, Vocabulary subtest. The Vocabulary subtest of the WISC-III (Wechsler, 1991) requires the participant to orally define stimulus words presented by the examiner. The words become increasingly difficult as the test progresses, and testing is stopped when the child misses four items in a row. Participants' responses in this study were scored two ways: the standard scoring method of 2-, 1-, or 0-point responses depending on the quality of the definition and a modified method in which a correct response received 1 point and an incorrect response received 0 points. The modified scoring method was used to reduce the influence of quality of word knowledge on the scoring to obtain a purer measure of breadth of word knowledge. Internal consistency reliabilities of the standard and modified scoring methods in this sample were .80 and .78, respectively.

Multiple Meanings subtest of the LPT-R. The Multiple Meanings subtest of the LPT-R (Richard & Hanner, 1995) requires the participant to recognize and define words that are presented in different contexts. The target item is presented in three sentence contexts, and the participant must name an appropriate synonym or definition for the word in each context. For example, the participant is given the word *rose* and the following three sentences: "Ask *Rose* to call me," "The

sun *rose* over the mountains,” and “She pricked her finger on the *rose*.” The internal consistency reliability of the Multiple Meanings test in this sample was .70.

Attributes subtest of the LPT–R. The Attributes subtest of the LPT–R (Richard & Hanner, 1995) measures the participant’s ability to name different attributes of a noun. The participant is asked to describe a noun, and scores are determined by the number of specific attributes named. Examples of possible attributes are function, components, color, accessories, size, shape, category, composition, and location. The internal consistency reliability of the Attributes test in this sample was .89.

WUF subtest of DIBELS. The WUF subtest of DIBELS (Good & Kaminski, 2002) requires participants to use grade-level target words in sentences. After the child speaks the sentence, the examiner presents another target word, and this continues for 1 min. The child’s score is the number of words correctly used in sentences during the 1-min testing period. Alternate-form reliabilities range from .59 to .71, depending on grade and length of time between administrations (Kaminski et al., 2004). Pilot testing with 40 third-grade students indicated an alternate-form reliability of .82 after using the Spearman–Brown prophecy formula to correct for test length. This correction was used because the participants’ scores for this study were determined by their cumulative performance on three forms of this test. When reliability was recalculated based on the forms given to all the students, the result was also .82.

Semantic Category Fluency. This measure was designed specifically for this study. Participants were required to name as many items as possible from eight categories, with 10 sec allowed for responding to each category. The categories administered to the participants were farm animals, fruits, clothes you wear, items of furniture, things that take you places (i.e., ways to travel), things used at school, wild animals, and things people drink. Scores were the total number of correct items named for all categories. Alternate-form reliability of the semantic category fluency task in this sample was .87.

Reading Comprehension Measures

The Florida Comprehensive Assessment Test–Sunshine State Standards (FCAT–SSS; Florida Department of Education, 2000, 2002a, 2005) and the Stanford Achievement Test–Ninth Edition (SAT–9; Harcourt Assessment, 1996) were used to assess reading comprehension. Scores from the two measures of reading comprehension were obtained from the school district for each participant. Although the two measures have similar response formats, the passages are constructed differently. The FCAT passages are primarily selected from previously

published material, whereas the SAT–9 passages are written specifically for the test. The quantity of literary and informational passages is similar for the SAT–9, whereas the FCAT generally consists of more literary passages at third grade. Finally, the FCAT passages tend to be longer than the SAT–9 passages.

FCAT–SSS Reading. The FCAT–SSS is a standardized, criterion-referenced test that is administered every spring to Florida public school students in Grades 3 through 10. On the reading section, third-grade students read passages and answer multiple-choice questions that evaluate their ability to comprehend written passages. Scores for FCAT Reading are reported on a scale of 100 to 500, with a mean score of 300 and standard deviation of 50 (Florida Department of Education, 2001). Reliability of the reading section of the FCAT ranges from .86 to .91 (Florida Department of Education, 2002b).

SAT–9 Reading Comprehension Test. The SAT–9 is a nationally norm-referenced, standardized test published by Harcourt Educational Measurement. It is administered every spring to Florida public school students in Grades 3 through 10. The reading portion of the SAT–9 assesses reading comprehension skills through the administration of written passages and multiple-choice questions. Scale scores for SAT–9 are reported on a scale of 527 to 817 (Florida Department of Education, 2003).

Procedure

The six measures of word knowledge were administered to each participant individually by a trained undergraduate student or the first author. Student testers were trained on the measures of word knowledge for approximately 8 hr by the first author. Additionally, they were required to demonstrate proficient administration of the measures before testing the participants.

Testing of each participant lasted approximately 60 min and was completed in two sessions, with one measure from each of the three hypothesized dimensions administered during each session. During the first testing session, the PPVT–III, the Multiple Meanings subtest of the LPT–R, and the Semantic Category Fluency test were administered. The Vocabulary subtest of the WISC–III, the Attributes subtest of the LPT–R, and the WUF subtest of DIBELS were administered during the second testing session. The measures were administered in this order to limit error variance due to time sampling.

RESULTS

Data Issues and Descriptive Statistics

The data set was examined for missing data, outliers, skewness, and kurtosis. Three data points were missing. Median imputation was used to handle missing data because less than 5% of the data were missing and the data set is large (Tabachnick & Fidell, 2001). Sixteen univariate outliers (4 for the PPVT–III, 1 for the WISC–III, 4 for Multiple Meanings, 1 for Attributes, 1 for WUF, and 5 for FCAT–SSS) were identified and recoded to be no greater than or less than two interquartile ranges from the median. 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 ($p < .001$) for a case being an outlier was used (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 are reported in Table 1. In all subsequent analyses, the modified version of the scoring system for

TABLE 1
Descriptive Statistics and Intercorrelations for All Observed Variables

Variable	1	2	3	4	5	6	7	8	9
1. PPVT–III	—								
2. WISC–III Vocabulary (standard)	.751	—							
3. WISC–III Vocabulary (modified)	.754	.968	—						
4. Multiple Meanings	.491	.601	.578	—					
5. Attributes	.272	.431	.398	.369	—				
6. WUF	.428	.490	.494	.384	.304	—			
7. Category Fluency	.356	.497	.483	.380	.276	.341	—		
8. FCAT–SSS Reading	.539	.613	.585	.479	.357	.291	.378	—	
9. SAT–9 Reading	.632	.711	.695	.505	.331	.289	.385	.776	—
<i>M</i>	132.670	27.750	15.374	8.887	39.700	43.188	4.212	354.207	669.429
<i>SD</i>	14.926	6.349	3.081	2.237	9.923	11.515	.740	52.359	38.521
Reliability	.942 ^a	.804 ^a	.781 ^a	.689 ^a	.887 ^a	.816 ^b	.873 ^b	—	—
Maximum possible	204	60	30	12	96	—	—	500	817

Note. $N = 203$. All correlations significant at $p < .01$. PPVT–III = Peabody Picture Vocabulary Test–Third Edition; WISC–III = Wechsler Intelligence Scale for Children–Third Edition; WUF = Word Use Fluency subtest; FCAT–SSS = Florida Comprehensive Assessment Test–Sunshine State Standards; SAT–9 = Stanford Achievement Test–Ninth Edition.

^aInternal consistency reliability. ^bReliability after correcting for test length.

the WISC Vocabulary test was used for two reasons. First, the modified system was more aligned conceptually with the definition of breadth used in this study. Second, the patterns of correlations between the modified scores and the other measures of vocabulary are consistent with the idea that it provided a slightly better measure of breadth. For example, although the differences are not significant, the PPVT-III was more highly correlated with the modified scoring system than with the standard scoring system. In addition, the two measures of depth, Multiple Meanings and Attributes, were more strongly related to the standard scoring system than to the modified scoring system.

Evaluation of CFAs

CFAs were conducted using AMOS 5.0 (Arbuckle, 2003) to evaluate measurement models. All CFAs were based on covariance matrixes. A CFA was initially used to confirm the fit of the data to the proposed three-factor model of Breadth, Depth, and Fluency. Additional nested models were tested to determine if any two-factor model or a one-factor model provided a better fit to the data than did the three-factor model. All models were identified, indicating that a unique set of parameter estimates was obtained. Fit indexes for the different models are presented in Table 2. The p value associated with the chi-square test is the probability of obtaining the observed variance-covariance matrix if the model is true for the population. Thus, a nonsignificant chi-square is desired. However, the chi-square test is overly sensitive to sample size (Bentler & Bonett, 1980) and makes this fit index less appropriate for evaluating model fit in this sample. Following the recommendations of Hu and Bentler (1999), a cutoff value of .95 was used as an indicator of good fit for the Tucker-Lewis Index and Comparative Fit Index. A cutoff value of .08 for the standardized root mean squared residual and .06 for the root mean squared error of approximation were also used as indicators of good fit. Multiple fit indexes are reported because each has both more and less desirable features. For example, the Tucker-Lewis Index is an incremental fit index that has the desirable feature of rewarding parsimonious models compared to complex models with many parameters. However, it is a sample-based fit statistic that can fall outside of the range of 0 to 1 and can produce values smaller than those of other fit indexes for small samples. The Comparative Fit Index is another incremental fit index that does not produce values outside the range of 0 to 1 but does not take parsimony into account (Hu & Bentler, 1998; Kline, 2005; Marsh & Balla, 1994; Marsh, Balla, & Hau, 1996). The fit indexes indicate that all models provided a good fit to the data.

Chi-square difference tests were conducted to compare the two-factor models and the one-factor model to the three-factor model. A significant chi-square difference test would indicate that the constraints imposed on the three-factor model to obtain the two-factor or one-factor models provide a worse fit to the data than does the three-factor model. Therefore, a nonsignificant chi-square difference test would in-

TABLE 2
Model Fit Indexes

<i>Model</i>	χ^2	<i>df</i>	<i>p</i>	χ^2/df	<i>RMSEA</i>	<i>SRMR</i>	<i>TLI</i>	<i>CFI</i>	$\chi^2_{difference}$ ^a
Three-Factor Model: Breadth, Depth, Fluency	4.23	6	.65	.70	.00	.02	1.00	1.00	
Two-Factor Model: Fluency, Breadth/Depth	14.52	8	.07	1.82	.06	.04	.97	.98	(2 vs. 1), 10.30**
Two-Factor Model: Breadth, Depth/Fluency	4.98	8	.76	.62	.00	.02	1.00	1.00	(3 vs. 1), .76
Two-Factor Model: Depth, Breadth/Fluency	11.79	8	.16	1.47	.05	.04	.98	.99	(4 vs. 1), 7.56*
One-Factor Model: General Word Knowledge	15.99	9	.07	1.78	.06	.04	.97	.98	(5 vs. 1), 11.76**

Note. *N* = 203. RMSEA = root mean squared error of approximation; SRMR = standardized root mean squared residual; TLI = Tucker–Lewis Index; CFI = Comparative Fit Index.

^aNumbers in parentheses indicate which models are being compared for chi-square difference tests.

p* < .05. *p* < .01.

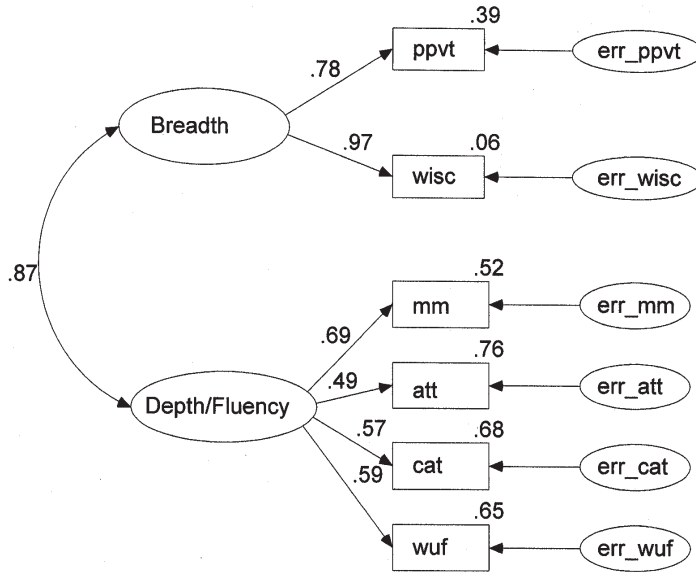


FIGURE 1 Two-factor measurement model of Breadth and Depth/Fluency. All regression weights and covariances are significant ($p < .001$). ppvt = Peabody Picture Vocabulary Test—Third Edition; wisc = Vocabulary subtest of Wechsler Intelligence Scale for Children—Third Edition; mm = Multiple Meanings subtest of Language Processing Test—Revised; att = Attributes subtest of Language Processing Test—Revised; cat = semantic category fluency; wuf = Word Use Fluency subtest of Dynamic Indicators of Basic Early Literacy Skills.

dicate that the two-factor or one-factor models provide a more parsimonious fit than does the three-factor model. These results are presented in Table 2. The two-factor model of Breadth and Depth/Fluency provided the best fit to the data. The two-factor model is illustrated in Figure 1 with standardized regression weights and error variances. Further results are presented later for the two-factor model.

Evaluation of Structural Model

Structural equation modeling was used to examine the predictive relations between the factors of word knowledge and reading comprehension. 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 two-factor model is presented in Figure 2. The model is identified.

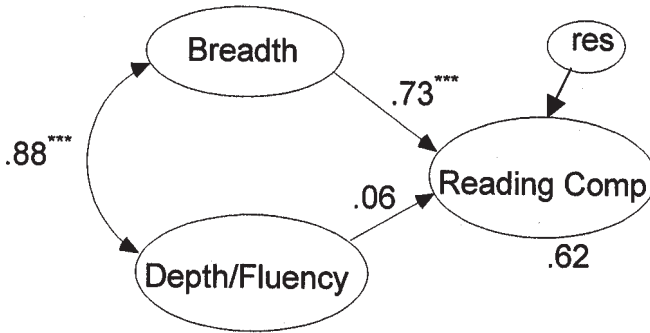


FIGURE 2 Two-factor structural model. *** $p < .001$

For the two-factor model, 62% of the variance in Reading Comprehension is explained by the Breadth and Depth/Fluency factors. Analysis of the regression weights indicates that Reading Comprehension increased .73 for each unit increase in Breadth, after controlling for Depth/Fluency. Reading Comprehension increased .06 for each unit increase in Depth/Fluency, after controlling for Breadth. However, the path between Depth/Fluency and Reading Comprehension was not significant.¹

Evaluation of Regression Analyses With Unit-Weighted Composites

To provide an examination of the relationships among these constructs using data that include typical measurement error, unit-weighted composites were created for each participant by converting their scores on each measure into z scores and then combining the z scores for each dimension of breadth, depth/fluency, and reading comprehension. Hierarchical regression analyses were conducted with the unit-weighted composites to determine if each dimension of word knowledge contributed unique variance in the prediction of reading comprehension beyond that

¹The relationship between each factor of word knowledge and Reading Comprehension was also examined in a confirmatory factor analysis. The benefit of utilizing a confirmatory factor analysis is that the relationships between the latent variables are considered separately and the analysis allows for the examination of the simple correlations between the latent variables. Results of the confirmatory factor analysis reveal correlations of .79 and .71 between the Reading Comprehension factor and the Breadth and Depth/Fluency factors, respectively. Although the reader may initially think that Depth/Fluency is not contributing to the prediction of Reading Comprehension by simple inspection of the structural model, the confirmatory factor analysis reveals that both factors are highly correlated with Reading Comprehension. Taken together, the two analyses reveal that Depth/Fluency cannot make much independent contribution because it is so highly correlated with Breadth.

TABLE 3
 Unique Variance Accounted for by Each Factor of Word Knowledge in the
 Prediction of Reading Comprehension

<i>Variable</i>	<i>Total R²</i>	<i>R² Change</i>	<i>Beta</i>	<i>F Change</i>
Breadth	.50	.19	.57	74.35**
Depth/Fluency	.50	.02	.19	7.75*

* $p < .01$. ** $p < .001$.

accounted for by the other dimension of word knowledge. The variable that is entered last in the analysis is the variable of interest, because one can determine whether it contributes unique variance in the prediction of reading comprehension by examining the R^2 change for significance once the other variables are in the equation. Table 3 displays the unique contribution of breadth and depth/fluency to the prediction of reading comprehension. Fifty percent of the variance in reading comprehension is accounted for by the two dimensions of word knowledge. Expressed in terms of a commonality analysis, breadth and depth/fluency explain 19% and 2% of the unique variance in reading comprehension, respectively. Twenty-nine percent of the variance in reading comprehension is explained by common variance. Correlations between the unit-weighted composites were calculated to examine the relationships between word knowledge and reading comprehension without both variables in the equation. Correlations of .70 and .56 were obtained between the reading comprehension composite and the breadth and depth/fluency composites, respectively.

DISCUSSION

In this study, three questions were addressed regarding the relationship between measures of word knowledge and reading comprehension in third-grade children. The first question asked if the three proposed dimensions of word knowledge (breadth, depth, and fluency) are distinguishable. The second question asked what the strength of the relationships is between the dimensions of word knowledge. Finally, the third question asked what the relationships are between the dimensions of word knowledge and reading comprehension. The answers to these three questions are addressed first in this section, followed by a discussion of their implications as well as their most important limitations. We conclude with a discussion of questions for the future.

This study showed that a two-factor model of Breadth and Depth/Fluency provided the best fit to the sample data. These results provide evidence that the three dimensions of word knowledge are not completely distinguishable from one an-

other using the tests employed in this study. One clear possibility is that depth and fluency are influenced by similar types of experiences with words, so they become empirically indistinguishable from one another. This is consistent with the work of Isabel Beck and her colleagues (Beck et al., 2002) who hypothesized that speed of access to word meaning improves as the word meaning is reinforced and flexibility of word knowledge is obtained. If having knowledge of multiple meanings for words implies broader and more frequent experience in identifying the meaning of the word, then it would make sense that depth of knowledge about word meaning would be highly associated with fluency in accessing a word's meaning.

This study also examined the relationships between the different dimensions of word knowledge. Results of the CFA and structural equation model revealed high correlations between the two factors of word knowledge, indicating that although the dimensions of breadth and depth/fluency are distinguishable from one another, they are highly related. This relationship, again, makes conceptual sense if we assume that vocabulary growth is stimulated by broad and deep experiences with words. The same general types of experiences that expose students to the meaning of many different words would also likely expose them to multiple meanings for individual words. And, if fluency of access to word meanings increases as a result of the number of times and the number of situations in which meaning is accessed, then fluency of access to word meanings should be highly correlated with both breadth and depth of vocabulary knowledge.

This study also evaluated the relationship between each dimension of word knowledge and reading comprehension to determine which dimension of word knowledge is most highly related to performance on measures of comprehension for text. These results provide evidence that breadth of word knowledge is most highly related to performance on a reading comprehension dimension composed of scores from the FCAT–SSS scoring system and the SAT–9 scoring system. Structural equation modeling revealed that reading comprehension increased the most as breadth increased. Regression analyses with unit-weighted composites indicated that breadth and depth/fluency accounted for significant amounts of unique variance in reading comprehension; however, the magnitude of the variance unique to depth/fluency was small. Further, over half of the variance in reading comprehension that was explained by the vocabulary measures was variance that the two vocabulary factors had in common.

The correlations between individual measures of word knowledge and reading comprehension obtained in this study (see Table 1) allow for direct comparisons with values obtained in previous research. Research with similar age students typically employed measures of breadth as indicators of word knowledge. The average correlation from a selection of these studies was .46, which is lower than the correlations obtained in this study between individual measures of breadth and reading comprehension. Results of hierarchical regression analyses with unit-weighted composites also allow a comparison between the results of this study and those ob-

tained from earlier examinations of the relationship between vocabulary and reading comprehension. The regressions with unit-weighted composites found a correlation of .70 between breadth and reading comprehension, which is again substantially higher than the average correlation from previous studies. In general, the results of this study demonstrate a stronger relationship between vocabulary and reading comprehension than does previous research.

Although the correlations obtained in this study are higher than most studies that have examined the relationship between measures of word knowledge and reading comprehension, one study presented a remarkably high correlation between vocabulary and reading comprehension. Davis (1942) reported a correlation of .86 between a measure of word knowledge and reading comprehension. This value is surprisingly high considering that measurement error was not extracted. The correlation from Davis's study is most likely higher than the values obtained in this study because Davis used adult participants. We argued earlier that the strength of the relationship between vocabulary and reading comprehension increases with age. Davis also used a measure of reading vocabulary. We also suggested earlier that reading vocabulary measures correlate higher with reading comprehension than do oral vocabulary measures, because reading vocabulary measures word reading ability in addition to knowledge of word meanings.

This is the first study to directly examine the relationships between reading comprehension and breadth, depth, and fluency of word knowledge using latent variables. An advantage of using latent variables is that the relationships among the variables are examined free of measurement error and reliability of the measures does not affect the relationships. Thus, the results from this study may provide a more accurate estimate of the relationships between the dimensions of word knowledge and reading comprehension than has been available from previous research.

The most important practical implication of this study is that a complete assessment of word knowledge in its relationship to reading comprehension should involve measures of breadth and depth/fluency. Although breadth is more robust than depth/fluency in its relationship to reading comprehension, the unit-weighted analysis indicated that depth/fluency does account for additional variance in reading comprehension under measurement conditions that are likely to be obtained in most educational settings. Although it is tempting to draw instructional implications from these findings, we are hesitant to do so because these analyses all involved concurrent correlations.

The concurrent nature of the relationships in this study is, in fact, one of the major limitations. Because the data were not collected in a longitudinal manner, it is inappropriate to make causal claims from the structural model. Many equivalent models with different causal paths could fit the data equally well (cf. MacCallum, Wegener, Uchino, & Fabrigar, 1993). One focus of future research should include longitudinal data collection so that causal statements are more appropriate.

We argued earlier that depth and fluency may be empirically indistinguishable from one another. However, another plausible explanation for the results obtained in this study is that depth and fluency were not adequately measured as separate constructs. This concern involves questions about the construct validity of the fluency and depth measures. Perhaps the measures chosen for depth and fluency did not measure these constructs as well as other measures that might be developed. Modification indexes were utilized in the CFA to determine if model fit could be improved by adding paths. These indexes revealed that model fit would improve if a path was added between WUF and SAT-9. Model fit would also improve, but to a lesser degree, if a path was added between Attributes and FCAT-SSS. These results seem to imply that these two measures of word knowledge may be assessing other constructs that are related to reading comprehension but that are different than those assessed by their companion measures. A question for future research is whether other measures of depth and fluency would allow these dimensions to be separately identified within a three-factor model of word knowledge.

Another possibility for further research might involve the construction of a task that assesses breadth, depth, and fluency of word knowledge using the same words. Perhaps a confrontational naming task could be used to assess the range of words that are least familiar to students, while rate of initial correct response serves as an indicator of fluency. Depth of word knowledge could then be obtained by asking for synonyms or antonyms of the word.

ACKNOWLEDGMENTS

This study was supported by a grant from the Florida Center for Reading Research and was the master's thesis of Kendra R. Tannenbaum. Richard K. Wagner was supported by Grant R305G030104 from the Institute of Education Sciences.

We thank Christopher J. Lonigan and Christopher Schatschneider for their helpful comments on this work. We are also grateful to the teachers, parents, children, and testers who participated in this study.

REFERENCES

- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.
- Arbuckle, J. L. (2003). *Amos 5.0 Update to the Amos Users Guide*. Chicago: SmallWaters Corporation.
- Beck, I. L., McCaslin, E., & McKeown, M. G. (1980). *The rationale and design of a program to teach vocabulary to fourth-grade students*. Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.
- Beck, I. L., & McKeown, M. G. (1983). Learning words well—A program to enhance vocabulary and comprehension. *The Reading Teacher*, 36, 622-625.

- Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.
- Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Davis, F. B. (1942). Two new measures of reading ability. *Journal of Educational Psychology*, *33*, 365–372.
- Davis, F. B. (1944). Fundamental factors of comprehension in reading. *Psychometrika*, *9*, 185–197.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test—Third Edition*. Circle Pines, MN: AGS Publishing.
- Florida Department of Education. (2000). *FCAT test item specifications*. Tallahassee: State of Florida, Department of State.
- Florida Department of Education. (2001). *FCAT briefing book*. Tallahassee: State of Florida, Department of State.
- Florida Department of Education. (2002a). *Lessons learned: FCAT Sunshine State Standards and instructional implications*. Tallahassee: State of Florida, Department of State.
- Florida Department of Education. (2002b). *Technical report: For operational test administration of the 2000 Florida Comprehensive Assessment Test*. Tallahassee: State of Florida, Department of State.
- Florida Department of Education. (2003). *Understanding FCAT reports*. Tallahassee: State of Florida, Department of State.
- Florida Department of Education. (2005). *The new FCAT NRT: Stanford Achievement Test Series, Tenth Edition*. Tallahassee: State of Florida, Department of State.
- Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills*. Eugene, OR: Institute for the Development of Educational Achievement.
- Hafner, L. E., Weaver, W. W., & Powell, K. (1970). Psychological and perceptual correlates of reading achievement among fourth graders. *Journal of Reading Behavior*, *2*, 281–290.
- Harcourt Assessment. (1996). *Stanford 9*. San Antonio, TX: Author.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
- Jenkins, J. R., Pany, D., & Schreck, J. (1978). *Vocabulary and reading comprehension: Instructional effects* (Tech. Rep. No. 100). Urbana: University of Illinois, Center for the Study of Reading.
- Kaminski, R. A., Good, R. H., Shinn, M. R., Smith, S. R., Laimon, D., Shinn, M., et al. (2004). *Development and research on DIBELS Word Use Fluency measure for kindergarten through third grades* (Tech. Rep. No. 13). Eugene: University of Oregon.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185–199.
- Marks, C. B., Doctorow, J., & Wittrock, M. C. (1974). Word frequency and reading comprehension. *Journal of Educational Research*, *67*, 259–262.
- Marsh, H. W., & Balla, J. (1994). Goodness of fit in confirmatory factor analysis: The effects of sample size and model parsimony. *Quality & Quantity*, *28*, 185–217.
- Marsh, H. W., Balla, J. R., & Hau, K. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issue and techniques* (pp. 315–353). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior*, *15*, 3–18.

- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.
- Nitsch, K. E. (1978). *Structuring decontextualized forms of knowledge*. Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN.
- Pany, D., & Jenkins, J. R. (1977). *Learning word meanings: A comparison of instructional procedures and effects on measures of reading comprehension with learning disabled students* (Tech. Rep. No. 25). Urbana-Champaign: University of Illinois.
- Raphael, T. E., Myers, A. C., Tirre, W. C., Fritz, M., & Freebody, P. (1981). The effects of some known sources of reading difficulty of metacomprehension and comprehension. *Journal of Reading Behavior*, 13, 325–334.
- Richard, G. J., & Hanner, M. A. (1995). *Language Processing Test—Revised*. East Moline, IL: LinguSystems, Inc.
- Snow, C. (2002, August). *Oral language development and demographic change*. Presentation at the Institute for Statewide Literacy Initiatives, Harvard Graduate School of Education, Cambridge, MA.
- Spearritt, D. (1972). Identification of sub-skills of reading comprehension by maximum likelihood factor analysis. *Reading Research Quarterly*, 8, 92–111.
- Stahl, S. A. (1998). Four questions about vocabulary knowledge and reading and some answers. In C. R. Hynd, S. A. Stahl, M. Carr, & S. M. Glynn (Eds.), *Learning from text across conceptual domains* (pp. 73–94). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, 20, 51–68.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics, fourth edition*. Needham Heights, MA: Allyn & Bacon.
- Tuinman, J. J., & Brady, M. E. (1974). How does vocabulary account for variance on reading comprehension tests? A preliminary instructional analysis. In P. Nacke (Ed.), *Twenty-Third National Reading Conference yearbook* (pp. 176–184). Clemson, SC: The National Reading Conference.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. New York: Psychological Corporation.
- Wolf, M., Miller, L., & Donnelly, K. (2000). Retrieval, Automaticity, Vocabulary Elaboration, Orthography (RAVE-O): A comprehensive, fluency-based reading intervention program. *Journal of Learning Disabilities*, 33, 322–324.

Manuscript received August 1, 2005

Accepted March 9, 2006

Copyright of *Scientific Studies of Reading* is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.