SCIENTIFIC STUDIES OF READING, *11*(3), 165–197

Development of Lexical Mediation in the Relation Between Reading Comprehension and Word Reading Skills in Greek

Athanassios Protopapas
Institute for Language & Speech Processing, Athena

Georgios D. Sideridis, Angeliki Mouzaki, and Panagiotis G. Simos
University of Crete

This study focuses on the shared variance between reading comprehension and word-level reading skills in a population of 534 Greek children in Grades 2 through 4. The correlations between measures of word and pseudoword accuracy and fluency, on the one hand, and vocabulary and comprehension skills, on the other, were sizeable and stable or increasing with grade. However, the unique contribution of word reading to comprehension became negligible after vocabulary measures were entered in hierarchical regression analyses, particularly for higher grades, suggesting that any effects of decoding on comprehension may be mediated by the lexicon, consistent with lexical quality hypothesis. Structural modeling with latent variables revealed an invariant path across grades in which vocabulary was defined by its covariation with reading accuracy and fluency and affected comprehension directly. It is argued that skilled word reading influences comprehension by strengthening lexical representations, at least when phonological decoding can be relatively effortless.

The point of reading is to understand text. From this perspective, the most important measure of reading performance concerns comprehension. An influential recent report unequivocally stated that "[u]nderstanding how to improve reading comprehension outcomes … should be the primary motivating factor in any future

Correspondence should be sent to Athanassios Protopapas, Institute for Language & Speech Processing, Artemidos 6 & Epidavrou, GR-15125 Maroussi, Greece. E-mail: protopap@ilsp.gr

literacy research agenda" (RAND Reading Study Group, 2002, p. xi). Therefore, the relation and contribution of so-called "component skills" to the development of reading comprehension constitutes a major focus of reading research, for good theoretical as well as practical, educational reasons.

In this study we examine the relation between reading comprehension and word-level skills such as decoding and fluency, focusing in particular on the variance in comprehension that can be accounted for by such component skills. We show that for our sample of Greek schoolchildren in Grades 2 through 4, this variance is shared with orally assessed vocabulary measures; therefore, it cannot be conceptualized as a unique contribution to comprehension from print-specific processes. We relate these findings to two prominent views regarding the factors underlying reading skill, and we discuss theoretically important implications worth investigating in future studies.

DISSOCIATION OF COMPREHENSION FROM DECODING

An influential model that has guided reading research into comprehension processes is the "simple view of reading" (henceforth, "Simple View"; Gough & Tunmer, 1986) which states that reading comprehension is the product of a general verbal component and a print-related component. That is, reading comprehension can be decomposed into a linguistic skill that can be assessed with listening comprehension measures, and a decoding skill that can be assessed with measures of isolated word and nonword reading accuracy and fluency. Thus, in the early stages of reading competence, reading comprehension is expected to be strongly related to decoding, because the ability to read the words limits the ability to understand the text. During the development of reading skills, as decoding expertise is gradually gained, comprehension is expected to be less dependent on decoding and more strongly related to the general linguistic comprehension skills, which limit the development of reading comprehension (Hoover & Gough, 1990).

According to the Simple View, there are alternative routes to reading comprehension failure, because decoding and listening comprehension may constitute independent obstacles to reading comprehension performance. Consistent with this view, poor readers can be classified into "garden variety" (Gough & Tunmer, 1986; Stanovich, 1988) when their poor reading comprehension can be attributed to language and verbal intelligence measures, and "dyslexic," when their reading performance is constrained by poor word-level skills such as inaccurate and/or slow grapheme to phoneme decoding. A more recent related approach was seen in the two-dimensional conceptualization of the relation between dyslexia and specific language impairment (SLI). In this approach, phonological difficulties are related to dyslexia, a word-level impairment (cf. Lyon, Fletcher, & Barnes, 2002), whereas problems in other language domains may impede comprehension; and combined problems in both domains lead to SLI (Bishop & Snowling, 2004).

Several predictions arising from the Simple View were confirmed by Hoover and Gough (1990). More recently, de Jong and van der Leij (2002) found that word-decoding speed and linguistic comprehension influence the development of reading comprehension from first through third grade, and they interpreted the stability of this relation as supporting the Simple View. Crucially, the Simple View posits that print-dependent decoding skills contribute unique variance to reading performance beyond any common variance with oral language skills. Specifically, decoding and listening comprehension are hypothesized to make *independent* contributions to reading comprehension (Tunmer & Hoover, 1992, p. 179). In support of this "relative independence" hypothesis, Muter, Hulme, Snowling, and Stevenson (2004) found reading comprehension in the British early school grades to be predicted by vocabulary knowledge and grammatical skills after controlling for the effects of early word recognition. In Norwegian 9-year-olds, Hagtvet (2003) found an association between decoding and comprehension but "a more extensive impact from oral language to reading comprehension than what is mediated by decoding" (p. 527), concluding in favor of the Simple View (albeit a "weak version" thereof, following Chen & Vellutino, 1997).

Testing less skilled comprehenders in French second grade, Megherbi and Ehrlich (2005) reported "low efficiency" in listening comprehension, regardless of decoding skill. Conversely, Cain, Oakhill, and Bryant (2000) found that children with reading comprehension difficulties but age-appropriate word reading skills did not differ from good comprehenders in measures of phonological processing except in tasks making great demands on working memory. Subsequently, Oakhill, Cain, and Bryant (2003) found that reading comprehension and word reading performance are accounted for by dissociable sets of measures. Specifically, verbal IQ and vocabulary accounted for variance in comprehension, as did comprehension monitoring, text integration, working memory, and story structure knowledge. In contrast, vocabulary did not account for variance in single-word reading accuracy, whereas reading rate, word recognition, and phonological awareness did.

Following up with these children, Cain, Oakhill, and Bryant (2004) found that, by 11 years of age, the correlation between (Neale) word reading and (Neale) reading comprehension was very low. This was consistent with the view that comprehension is increasingly independent of word reading and that word reading and comprehension skills can be differentiated. Similarly, in the Greek seventh grade, Protopapas and Skaloumbakas (2007) found that a reading comprehension measure failed to distinguish between the general population and a clinical population of children referred for evaluation of learning disability. They also found that word and nonword reading accuracy, along with spelling accuracy and phoneme deletion, loaded on a different factor from the one on which reading comprehension loaded along with other measures of verbal, nonverbal, and mathematical ability.

This pattern of factor loadings, with comprehension lying along a different dimension from reading and spelling accuracy, has also been reported in U.S. college students (Perfetti & Hart, 2002). Similarly, in a large population of university students, Landi (2005) reported that high-level skills such as reading comprehension and vocabulary were partly dissociable from low-level reading skills such as decoding and spelling ability.

Recently, Braze, Tabor, Shankweiler, and Mencl (in press) tested the Simple View with a group of adolescents and young adults in continuing education with a wide range of reading abilities, including many who had struggled in primary and secondary education. Braze et al. argued that if variance in reading comprehension is exhausted by listening comprehension and word identification measures, there should be no additional contribution made by vocabulary measures. Their findings did not support this hypothesis and pointed instead to a significant role of vocabulary beyond the two components of the Simple View. An important aspect of these findings is that the contribution of vocabulary was much more substantial in predicting variance in reading comprehension than in listening comprehension. Braze et al. attributed this asymmetry to the impoverished nature of print, compared to speech, in its potential to activate lexical representations, which makes high-quality lexical representations more important for reading than for oral comprehension.

Overall, the diminishing role of word-level decoding skills for reading comprehension is consistent with the Simple View. Recent investigations have tested an additive instead of a multiplicative model (Chen & Vellutino, 1997; Joshi & Aaron, 2000; Savage, 2006). They have also revealed that speed measures can contribute to a fruitful extension of the model (Johnston & Kirby, 2006; Joshi & Aaron, 2000; but see Adlof, Catts, & Little, 2006), largely supporting the core conceptualization offered by the Simple View. However, the role of vocabulary measures, at the behavioral level, and of lexical representations, at the cognitive level, remains to be clarified. Because word forms are accessed by decoding in word identification, and word meanings are combined to derive text meaning for comprehension, it is plausible that lexical representations may constitute the link between the two levels of skill.

The Importance of Lexical Representations

The "lexical quality" hypothesis posits a special role for lexical representations in the development and expression of reading skills (Perfetti & Hart, 2001, 2002). The quality of lexical representations concerns the strength of associative connections between constituents of lexical specification at the semantic, phonological, and orthographic levels. Specifically, "a lexical representation has high quality to the extent that it has a fully specified orthographic representation (a spelling) and redundant phonological representations (one from spoken language and one recoverable from orthographic-to-phonological mappings)" (Perfetti & Hart, 2001, p. 68).

Lexical representations primarily concern word meanings; that is, semantic representations forming associative networks. However, the semantic specification must be connected to its corresponding phonological and orthographic forms, via which it is accessed in oral and written language. Thus, lexical quality is about "detailed knowledge about word forms and meanings" that can "drive rapid processing" (Perfetti & Hart, 2001, p. 76). High-quality lexical representations are strongly activated by their constituent representations and facilitate access to their contents. The origin of high-quality representations may be sought in the amount of experience with both oral and written language.

In support of the lexical quality hypothesis, Beck, Perfetti, and McKeown (1982) taught new words to fourth graders and found large performance improvements on a number of semantic tasks, despite imperfect learning of the novel vocabulary items. Most important, improvements seemed to extend to untrained material, suggesting that the children were acquiring vocabulary skills and not merely new words.

An important aspect of evaluating the connection between lexical quality and the Simple View concerns the possible relations between word-level reading skills, such as decoding accuracy and fluency, on the one hand, and reading comprehension and other higher level verbal skills, on the other hand. The lexical quality approach tends to emphasize the association between decoding and comprehension, whereas the Simple View approach seems to focus on the dissociation. Given the substantial shared variance among measures of reading comprehension and component skills, it is unclear whether there is a direct opposition between the two approaches or whether they concern different aspects of skilled reading and of learning to read. One possibility is that lexical quality contributes to reading comprehension performance independently of the decoding and listening comprehension factors of the Simple View, accounting for additional unique variance (cf. Braze et al., in press). Another possibility is that lexical quality mediates, at least in part, the development and expression of both listening comprehension and word-level decoding skills, and so its contribution to reading comprehension would be difficult to disentangle from that of the two factors of the Simple View.

In this context, it is necessary to identify valid indexes of lexical quality. Standard tests of expressive and receptive vocabulary seem to provide a reasonable starting point. Perfetti and Hart (2001) considered vocabulary measures as properly concerning only the semantic component of lexical representations, leaving out the phonological and orthographic components. However, it may be argued that a large and strong vocabulary is incidentally related to overall high lexical quality by virtue of how the vocabulary is presumably established and refined in the first place; that is, primarily by reading, which strengthens the orthographic and phonological connections to the semantic networks.

The Role of Vocabulary Measures

In a recent review, Joshi (2005) concluded that "[d]eveloping a larger vocabulary is often a critical factor in improving reading comprehension" (p. 215) and noted that insufficient attention has been paid to the role of vocabulary knowledge in fluent reading. The connection between vocabulary measures and reading comprehension may lie in the effect of reading practice. In a study aiming to identify the strongest independent predictors of reading comprehension among children in Grades 3 through 5, Goff, Pratt, and Ong (2005) found that irregular word reading left no variance in reading comprehension to be accounted for by nonword reading. They suggested that decoding is no longer important once lexical skills have been accounted for. In their sample, a receptive vocabulary measure (the Peabody Picture Vocabulary Test–III; PPVT–III) was significantly related to exposure to print and was more strongly related to irregular word reading than to nonword reading. Assuming that children with better vocabulary tend to read more, and are more skilled at reading irregular words, Goff et al. suggested that irregular word reading ability captures variance in both word identification skills and general language ability.

The potential for reciprocal interactions between lexical skills, word-level reading ability, and comprehension was highlighted in a longitudinal study of reading development by Nation and Snowling (2004). In their study, both expressive vocabulary and listening comprehension were found to be better longitudinal predictors of word recognition than a "semantic composite" that included semantic fluency and a synonym judgment task. Nation and Snowling favored a connectionist conceptualization of the reading process "in which to consider the interplay of phonological and meaning-based factors in word recognition development" (p. 353).

Seigneuric and Ehrlich (2005) regressed reading comprehension onto decoding, vocabulary, and working memory in Grades 1, 2, and 3, and they found significant independent contributions of decoding (decreasing with grade) and vocabulary (increasing with grade) in every grade (controlling for age). Similarly, Yovanoff, Duesbery, Alonzo, and Tindal (2005) tested a multiple regression model predicting reading comprehension and found the importance of vocabulary increasing with grade, although that of fluency decreased. The importance of vocabulary in the longitudinal prediction of reading comprehension was also noted by Bast and Reitsma (1998), who considered their measure of receptive vocabulary as an index of general language ability, and thus, interpreted their findings as largely consistent with the Simple View.

It is notable that vocabulary measures are typically considered indicative of language ability and not of word recognition skill, and in the context of the Simple View they would be expected to contribute to the "language" component and not to the "decoding" component. Because of this, the effects of word-level decoding, and associated fluency, are thought of as primary, hence direct, and thus far the effects of vocabulary are typically examined after decoding has already been partialled out of the comprehension variance. This approach may obscure potential direct links between vocabulary and comprehension, as would be expected by the lexical quality hypothesis insofar as vocabulary measures indeed assess the quality of lexical representations. Moreover, because only a single measure of vocabulary has been usually employed—either expressive or receptive—it is not known whether all relevant variance is fully exploited or perhaps the "vocabulary" construct would benefit from additional, complementary measures.

In this article we present cross-sectional data from Greek children in the second through fourth grades, focusing on the relations between vocabulary measures, word and nonword reading accuracy and fluency, and reading comprehension performance. Taking vocabulary as an index of lexical skill and, by extension, of lexical representation quality, we suggest that lexical representations may mediate the relation between word- and text-level performance. If that is the case, then vocabulary measures should not only correlate strongly with both word-level reading and text comprehension but should also take up most of the common variance between these two levels of reading performance. A potential connection is thus suggested between the lexical quality hypothesis and the Simple View in an effort to understand both the observed dissociations and the established associations between levels and skills.

METHOD

Participants

The sample included 534 children from elementary school Grades 2, 3, and 4, from 17 Greek elementary schools in Crete, Attica (including the Athens metropolitan area) and the Ionian islands. School selection followed a stratified randomized approach in an effort to include units representative of urban (seven schools), rural (three schools), and semi-urban areas (seven schools). Children were selected randomly from each class, but only those children whose parents gave written permission to participate in the research were included in the study. All participating students were fluent speakers of the Greek language, had never been retained in the same grade, and did not suffer from any mental or emotional impairment that prohibited their enrollment in the regular education class of their school. All children were tested individually in two 40-min sessions by a group of undergraduate and graduate students during a period of 3 weeks in March of 2005. Examiners had undergone long and rigorous training and were closely monitored by the study coordinators to standardize administration procedures. Table 1 shows the numbers of boys and girls and the mean age for each grade.

TABLE 1
Sex and Age (In Years;Months) of Participants for Each Grade

Grade	*Boys*	*Girls*	*Age (SD)*
2	92	96	7;9 (0;3)
3	83	93	8;9 (0;3)
4	79	91	9;9 (0;4)

Assessments

Children were tested on word and pseudoword reading accuracy, pseudoword and sight word efficiency (fluency), reading comprehension, rapid automatized naming (RAN), spelling, a visual-constructive task, and expressive and receptive vocabulary.

Word reading fluency. A sheet with 112 isolated words printed in four columns was presented for the child to read as quickly as possible. Words ranged in length from one to six syllables and were presented in order of increasing length. Each child was instructed to name each word as fast as possible without making errors starting from the top of each column and moving to the bottom in 45 sec. Children received one point for each item that they accurately named (correct phonological decoding and stress) within this time limit.

This test was designed to assess the efficiency of automatic recognition of high frequency words. Words were initially selected on the basis of frequency of appearance within the Hellenic National Corpus (Hatzigeorgiu et al., 2000; hnc.ilsp.gr), a corpus of approximately 34 million lexical units compiled from a wide selection of texts (mainly popular Greek books published after 1990 and daily newspapers). All 112 tokens in the word list were among the 1,000 most frequent words in the corpus. To further ensure that a sufficient number of words visually familiar to the youngest students in the study was included in the list, 30 items were among those appearing in the basic vocabulary selection of the second grade reading textbook.

Word reading accuracy. Subtest 5 of the Test of Reading Performance (TORP; Padeliadu & Sideridis, 2000; Sideridis & Padeliadu, 2000) was used. A sheet with 40 isolated words printed in two columns in order of ascending difficulty was presented for the child to read without time pressure. Words ranged in length from two to five syllables. Responses were scored with 0 (*inaccurate item reading*), 1 (*phonologically correct but inaccurate use of stress*), or 2 (*phonologically accurate and correctly stressed response*). Administration was discontinued when students scored zero on six consecutive items.

Pseudoword reading fluency. A sheet with 70 isolated pseudowords printed in four columns was presented for the child to read as quickly as possible. Pseudowords ranged in length from one to six syllables and were presented in order of increasing length. Each child was instructed to name each word as fast as possible without making errors starting from the top of each column and moving to the bottom in 45 sec. Children received one point for each item that they accurately named (correct phonological decoding and stress) within this time limit.

This test was designed to assess the efficiency of speeded pseudoword decoding. Pseudowords were constructed by altering one or two letters in 70 words matched on mean frequency of appearance with those included in the word list, maintaining some of the phonological or morphological (or both) characteristics of the original (high frequency) word.

Pseudoword reading accuracy. Subtest 6 of the TORP (Padeliadu & Sideridis, 2000; Sideridis & Padeliadu, 2000) was used. A sheet with 19 isolated pseudowords printed in order of ascending difficulty was presented for the child to read without time pressure. Pseudowords ranged in length from two to three syllables. Responses were scored with 0 (*inaccurate item reading*), 1 (*phonologically correct but inaccurate use of stress*), or 2 (*phonologically accurate and correctly stressed response*). Administration was discontinued when students scored zero on six consecutive items.

Reading comprehension. Reading comprehension was assessed with Subtest 13 of the TORP (Padeliadu & Sideridis, 2000; Sideridis & Padeliadu, 2000). The test includes six passages of ascending length (word counts per passage: 19, 26, 51, 65, 97, and 85), each followed by two to four multiple choice questions (with four options each). Children were asked to read each passage aloud and then to read and answer all the questions following each passage. Passages and questions were presented in a test booklet and children were allowed to look at the passages while answering the questions. Passages (five narratives, one expository) became progressively more difficult by increasing vocabulary level and syntactic complexity. Most comprehension questions related to story characters and their actions, whereas a few of the later questions concerned story topic and main idea. The total number of questions for the six passages was 18 (13 explicit, answered with information found directly in the passage, and 5 implicit, involving some "higher" thinking in terms of reader judgment based on the text information) and each was scored with 0 (*inaccurate answer*) or 1 (*accurate answer*). Responses were scored during test administration to allow application of a floor-performance discontinuation criterion (when all questions following a passage were answered incorrectly), in which case questions to subsequent (not administered) passages were also scored with zero.

Spelling accuracy. Single-word spelling ability was assessed using a list of 60 words selected from the basic vocabulary selection in reading textbooks used in Grades 1 through 6. Selection of words ensured representation of key instructional units of grammar and spelling rules taught in each grade (i.e., verb past tense, noun clauses, etc.) and were arranged in ascending order of difficulty based on their grade-level appearance and on teacher ratings of their spelling difficulty. The examiner first pronounced each word in isolation and then in context to demonstrate its use. After repeating the word in isolation, the examiner asked the child to write the word on a numbered form. Each word was scored with one point for accurate spelling. Missing or misplaced stress diacritics were not scored due to their high frequency of occurrence. The spelling task was discontinued when the child scored zero on six consecutive items.

RAN of letters. A measure of RAN was administered, composed of five Greek letters (β, σ, κ, π, and α) presented 10 times each, arranged in 5 rows by 10 columns, following the standard form (Denckla & Rudel, 1974). Children were instructed to name each item as fast as they could without omitting any. The total time (in seconds) to name the entire set of items was noted and used in the analysis after direction inversion (to keep correlations positive).

Receptive vocabulary. The Peabody Picture Vocabulary Test–Revised (PPVT–R; L. M. Dunn & Dunn, 1981), adapted for Greek, was administered as an index of receptive lexical skills. The original PPVT–R is a widely used receptive vocabulary test of well-established and highly valued psychometric properties. For the Greek adaptation we used the original picture templates, but we altered either the order of appearance of some items/words or the items/target words in few templates. The alterations were made mainly to address differences in the difficulty that certain words pose in each language. The new item order was based on pilot data from 35 children who were tested with the original stimulus order.

In this task, each child was asked to identify one picture out of four that best represented the word given orally by the examiner. Test items/words were arranged in ascending order of difficulty and were administered in backward order until a base of 6 correct answers was established. Once the base of 6 correct responses had been established, children then received credit for all previous items. The test was discontinued when the child gave 8 incorrect answers within 10 consecutive questions. Raw scores (number of correct items) were used in the analyses because there is no standardization for the Greek population.

Expressive vocabulary. The vocabulary subtest of the WISC–III (Wechsler Intelligence Scales for Children, Greek standardization; Georgas, Paraskevopoulos, Bezevegis, & Giannitsas, 1997) was administered as a measure of expressive vocabulary. The vocabulary subtest includes 30 word items that are given to children for definition and are scored with 2, 1, or 0 points each, depending on word understanding and richness of expression. The subtest is discontinued after four consecutive zero-scoring responses. Participant answers were recorded verbatim on test protocols and were later rescored by the study coordinator and a small team of examiners to ensure high reliability. Standard scores were used in the analyses.

Nonverbal intelligence. The block design subtest of the WISC-III (Greek standardization; Georgas et al., 1997) was administered as a rough index of nonverbal ability. This subtest includes 12 designs that children are asked to recreate using a specified number (4–9) of two-colored blocks within a time limit. The subtest is discontinued after two consecutive failures. Child attempts are scored for the accuracy and speed of reproduction. Standard scores were used in the analyses.

Combined standard scores for the vocabulary and block design subtests provide an estimate of full scale IQ (Sattler, 1982).

RESULTS

The analyses reported next include all available data without any modification, except for an inverse transformation of the word reading accuracy (TORP–5) measure and a log transformation of the RAN measure, due to extreme skewness. No cases were removed or data points replaced. There were six missing data points in all, scattered over different variables and cases, and eight extreme multivariate outliers (detected via a robust principal component method at a χ^2 criterion of $p < .00001$). All analyses were repeated after removing the corresponding 14 cases, without affecting the pattern of significant findings or the conclusions; therefore, here we report the results from the full data set only.

Table 2 presents means and standard deviations, per grade, for each measure. The number of children reaching the stopping criterion during administration of the reading comprehension test is shown in Table 3. An index of reliability for each unstandardized measure with multiple items was calculated using Cronbach's alpha over the entire sample and was found to be .79 for word reading accuracy, .90 for nonword reading accuracy, .79 for reading comprehension, .95 for PPVT, and .94 for spelling. For word reading fluency, a second list was administered, and the correlation between the two lists was $r = .94$.

Multivariate Analysis

The correlations among the measures in each grade are shown in Table 4. As expected, most measures were intercorrelated, with the notable exception of RAN, which was not significantly correlated with comprehension or measures of verbal and nonverbal ability in any grade. Table 5 shows the partial correlations of reading

TABLE 2
Basic Descriptive Statistics Per Grade

		Grade 2[a]		Grade 3[b]		Grade 4[c]	
Measure	*Possible range*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age (months)	—	93.5	3.5	105.5	3.5	117.3	4.2
Word fluency	0–112	42.7	11.1	54.4	12.8	62.4	12.5
Pseudoword fluency	0–70	16.5	5.9	22.5	9.1	26.1	8.6
WISC block design	0–20	9.4	3.2	9.5	3.4	9.5	3.0
WISC vocabulary	0–20	9.6	3.0	9.8	2.6	9.5	3.0
RAN letters	—	30.8	6.1	26.3	5.5	23.5	5.3
Word accuracy	0–80	70.8	6.8	74.8	5.3	76.5	4.4
Pseudoword accuracy	0–38	25.8	6.3	29.9	6.8	31.3	6.5
PPVT	0–173	104.8	17.0	118.8	14.3	125.1	14.3
Spelling	0–60	23.4	6.8	32.8	9.0	38.9	10.2
Reading comprehension	0–18	8.9	3.7	10.7	3.3	11.7	3.1

Note. [a]$N = 188$. [b]$N = 176$. [c]$N = 170$.

TABLE 3
Number of Children Reaching, but Not Progressing Beyond, the Indicated Passage in the Reading Comprehension Task

	Passage			
Grade	*2*	*3*	*4*	*5*
2	6	18	42	66
3	1	5	14	43
4	0	5	9	33
All	7	28	65	142

Note. Out of Six Passages Total; the Stopping Criterion Is Not Applied at the First Passage.

comprehension with the individual measures, after controlling for vocabulary and nonverbal ability (i.e., three controlled variables: WISC block design, WISC vocabulary, and PPVT). Note that these partial correlations were substantially weakened and diminish with increasing age, dropping to nonsignificance by Grade 4.

In linear regression analyses, a substantial amount (up to 48%) of variance in reading comprehension could be accounted for by the available measures. However, the contribution of word-level skills, as indexed by accuracy and fluency measures, entered after block design and vocabulary measures, was very small and not statistically significant after Grade 3. Table 6 shows the results of the regressions for each grade, in which nonverbal ability and vocabulary were always entered in Steps 1 and 2, respectively, followed by accuracy and speed measures, in Steps 3 and 4. It is important to note that the unique contribution of accuracy

TABLE 4
Correlation Coefficients (Pearson's *r*) Among All Measures, Separately for Each Grade

	PWFL	*WBL*	*WVOC*	*RAN*	*WAC*	*PWAC*	*PPVT*	*SPEL*	*RCOMP*
Grade 2[a]									
WFL	.71***	.10	.29***	.51***	.54***	.31***	.16	.67***	.34***
PWFL	—	.05	.21*	.39***	.51***	.51***	.15	.51***	.31***
WBL		—	.24*	–.02	.27**	.19	.29***	.11	.27**
WVOC			—	.14	.28**	.19*	.53***	.35***	.47***
RAN				—	.26**	.20*	.01	.21*	.11
WAC					—	.50***	.27**	.51***	.39***
PWAC						—	.18	.31***	.19*
PPVT							—	.15	.46***
SPEL								—	.35***
Grade 3[b]									
WFL	.74***	.15	.29***	.64***	.53***	.50***	.23*	.76***	.25**
PWFL	—	.06	.21*	.39***	.61***	.61***	.17	.61***	.17
WBL		—	.32***	–.01	.17	.07	.37***	.16	.30***
WVOC			—	.17	.31***	.07	.55***	.41***	.40***
RAN				—	.23*	.28**	.13	.38***	.11
WAC					—	.57***	.32***	.58***	.30***
PWAC						—	.12	.54***	.12
PPVT							—	.33***	.50***
SPEL								—	.36***
Grade 4[c]									
WFL	.71***	.20*	.29**	.51***	.45***	.48***	.34***	.70***	.25**
PWFL	—	.15	.28**	.45***	.52***	.62***	.32***	.60***	.19
WBL		—	.39***	.07	.25*	.23*	.45***	.22*	.37***
WVOC			—	.21*	.30**	.28**	.58***	.35***	.49***
RAN				—	.20*	.28**	.18	.33***	.08
WAC					—	.58***	.45***	.58***	.34***
PWAC						—	.36***	.60***	.23*
PPVT							—	.46***	.66***
SPEL								—	.36***

Note: WFL = word fluency; PWFL = pseudoword fluency; WISC = Wechster Intelligence Scales for Children; WBL = WISC block design; WVOC = WISC vocabulary; RAN = rapid automatized naming of letters; WAC = word reading accuracy; PWAC = pseudoword reading accuracy; PPVT = Peabody Picture Vocabulary Test; SPEL = spelling accuracy; RCOMP = reading comprehension

[a]$N = 188$. [b]$N = 176$. [c]$N = 170$.

*$p < .01$. **$p < .001$. ***$p < .0001$.

measures to reading comprehension diminishes progressively with increasing grade, reinforcing the conclusion that the developmental trend does not reflect some idiosyncratic property of the particular sample. In contrast, the unique contribution of vocabulary, after taking into account all other measures, does not diminish and remains statistically significant in every grade (see Table 7).

TABLE 5
Partial Correlations of Word-level Measures with Reading Comprehension, Controlling for Wechster Intelligence Scales for Children III Block Design, Wechster Intelligence Scales for Children III Vocabulary, and Peabody Picture Vocabulary Test, Separately for Each Grade

	Grade		
Measure	*2*	*3*	*4*
Word fluency	.24*	.12	.01
Pseudoword fluency	.24*	.09	.05
RAN letters	.08	.05	.07
Word accuracy	.26**	.14	.05
Pseudoword accuracy	.07	.07	.04
Spelling	.25*	.21*	.07

Note. RAN = rapid automatized naming of letters.
$p < .01$. $^{**}p < .001$.

Table 8 shows the number of children at ceiling on the word and pseudoword reading accuracy tests. As expected for a language with a shallow orthography, such as Greek, a substantial number of children excel at these tests, increasingly so at higher grades. Could the reduced variance, caused by the ceiling effect, account for the diminished importance of word-level skills for comprehension? We believe not, for the following reasons. First, mean accuracy performance remains at least one standard deviation below ceiling (see Table 2). Second, word and pseudoword reading accuracy remain highly and significantly correlated with reading comprehension and with other relevant measures, such as fluency and spelling, in all three grades (see Table 4); the strength of the zero-order correlations hardly drops with increasing grade. Therefore the available variance in accuracy is at least sufficient for these relations to be clearly detected. Third, the zero-order and partial correlation results for word and pseudoword fluency, where there is no ceiling performance, strongly support the same conclusions regarding the relation of greatest interest here; that is, between reading comprehension and word-level "component" skills.[1]

Structural Analysis

To test the hypothesis that comprehension is a higher order reading skill that is predicted by lexical skills, which in turn are predicted by the students' word-level

[1]In languages with shallow orthography, where ceiling accuracy levels are not uncommon, fluency measures are considered better indicators of word-level skills and more useful for the detection of word-level reading disability (see, e.g., for Finnish: Holopainen, Ahoren, & Lyytinen, 2001; German: Landerl, 2001; Italian: Tressoldi, Stella, & Faggella, 2001; Spanish: Jiménez González & Hernández Valle, 2000). For Greek in particular, see Porpodas (1999) for beginning readers and Protopapas and Skaloumbakas (2007) for seventh-grade children.

TABLE 6
Results of Linear Regressions, Separately for Each Grade, with Reading Comprehension as the Dependent Variable and Vocabulary Entering Before Reading Skills

Step	*Variables*	R^2	ΔR^2	ΔF	df_1	df_2	*p*
Grade 2[a]							
1	WISC block design	.08	.08	15.88	1	183	.000
2	WISC vocabulary, PPVT	.30	.22	28.17	2	181	.000
3	Spelling, word accuracy, pseudoword accuracy	.37	.07	6.29	3	178	.000
4	RAN letters, word fluency, pseudoword fluency	.38	.01	0.86	3	175	.461
Grade 3[b]							
1	WISC block design	.09	.09	16.81	1	174	.000
2	WISC vocabulary, PPVT	.28	.19	23.02	2	172	.000
3	Spelling, word accuracy, pseudoword accuracy	.32	.04	3.07	3	169	.029
4	RAN letters, word fluency, pseudoword fluency	.32	.00	.15	3	166	.927
Grade 4[c]							
1	WISC block design	.13	.13	24.95	1	165	.000
2	WISC vocabulary, PPVT	.47	.33	5.79	2	163	.000
3	Spelling, word accuracy, pseudoword accuracy	.47	.01	0.72	3	160	.540
4	RAN letters, word fluency, pseudoword fluency	.48	.01	0.46	3	157	.712

Note. *WISC = Wechster Intelligence Scales for Children; PPVT = Peabody Picture Vocabulary Test; RAN = rapid automatized naming of letters.*
[a]*N* = 188. [b]*N* = 176. [c]*N* = 170.

reading skills (accuracy and fluency), latent-variable structural equation modeling was used. Initially a measurement model was tested with and without latent variable correlations to ensure that the latent constructs were defined properly. Evidence to that effect would be manifested if all indicator paths defining a construct were significant and only if they defined their hypothesized construct (no crossloadings were allowed). Following evaluation of a structural model employing the full sample, a series of latent variable models were tested to evaluate the hypotheses that (a) the slopes linking the constructs with each other are invariant across grades, and (b) the latent means are invariant across grades. All models were run using EQS 6.1 (Bentler, 2004).

Measurement model. A measurement model was set up in which indicator variables included three untimed accuracy measures for "accuracy" (word and

TABLE 7
Results of Linear Regressions, Separately for Each Grade, with Reading Comprehension as the Dependent Variable and Vocabulary Entering After Reading Skills

Step	*Variables*	R^2	ΔR^2	ΔF	df_1	df_2	*p*
Grade 2[a]							
1	WISC block design	.08	.08	15.88	1	183	.000
2	Spelling, word accuracy, pseudoword accuracy	.23	.15	11.26	3	180	.000
3	RAN letters, word fluency, pseudoword fluency	.24	.01	.84	3	177	.475
4	WISC vocabulary, PPVT	.38	.14	19.42	2	175	.000
Grade 3[b]							
1	WISC block design	.09	.09	16.81	1	174	.000
2	Spelling, word accuracy, pseudoword accuracy	.22	.13	9.29	3	171	.000
3	RAN letters, word fluency, pseudoword fluency	.22	.01	.40	3	168	.757
4	WISC vocabulary, PPVT	.32	.10	11.99	2	166	.000
Grade 4[c]							
1	WISC block design	.13	.13	24.95	1	165	.000
2	Spelling, word accuracy, pseudoword accuracy	.23	.10	7.21	3	162	.000
3	RAN letters, word fluency, pseudoword fluency	.24	.00	.18	3	159	.909
4	WISC vocabulary, PPVT	.48	.24	36.04	2	157	.000

[a]*N* = 188. [b]*N* = 176. [c]*N* = 170.

TABLE 8
Number of Children in Each Grade Exhibiting Ceiling Performance in Word Reading Accuracy (TORP–5 Score = 80) and Pseudoword Reading Accuracy (TORP–6 Score = 38).

Grade	*Word*	*Pseudoword*
2	7	6
3	30	22
4	46	24

Note. TORP–5 = Test of Reading Performance, Subtest 5; TORP–6 = Test of Reading Performance, Subtest 6.

pseudoword reading accuracy as well as spelling), three timed reading measures for “fluency” (two word fluency lists and pseudoword fluency), a receptive and an expressive measure for “vocabulary” (PPVT–R and WISC–III vocabulary), and six passages with associated questions for “comprehension.” A latent variable model in which all indicators loaded on their respective constructs suggested good

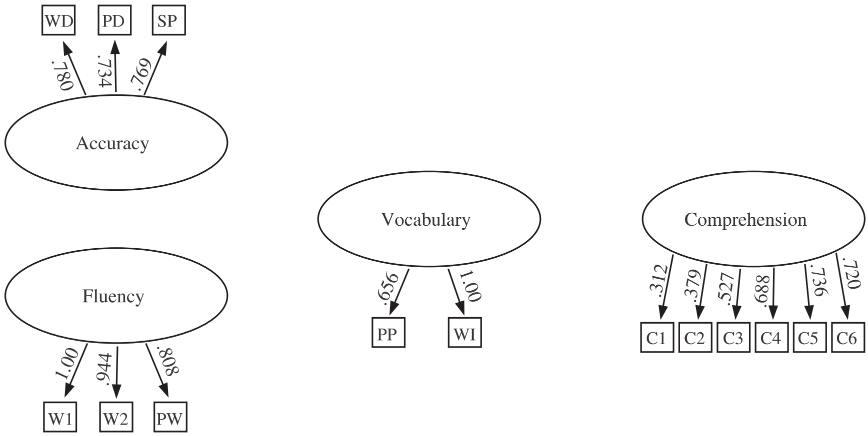

FIGURE 1 Measurement model in which indicators were linked to their respective latent estimates. The model does not include covariations between constructs. All measurement paths were significant and substantially larger than suggested minimum estimates (i.e., .30; Gorsuch, 1983).

construct validity, as all paths were statistically significant at the .05 level, $\chi^2(77, N = 534) = 1165.522$, $p < .001$ (Comparative Fit Index [CFI] = .737, Non Normative Fit Index [NNFI] = .690, Incremental Fit Index [IFI] = .738, Standardized Root Mean Squared Residual [SRMR] = .296, Root Mean Squared Error of Approximation [RMSEA] = .166; Model 1, Table 9). Thus, there was evidence that the reading constructs were defined well by their respective indicators.[2]

Structural model. The first structural model postulates that comprehension is a function of vocabulary and that vocabulary is defined by a covariation with fluency plus a covariation with accuracy. Additionally, direct paths link accuracy and fluency to comprehension, expected to lose strength as a function of grade level as students move from reliance on early processes to fluent understanding of text. This model provides acceptable fit to the data, although the chi-square was significant, $\chi^2(70, N = 534) = 239.336$, $p < .001$ (CFI = .959, NNFI = .947, IFI = .959,

[2]Potential misfits were explored to determine whether the model could be further improved before including structural parameters. Overall model fit in measurement models may appear poor because all factors are tested as orthogonal entities and the absence of modeled covariations reduces overall model fit (Dunn, Everitt, & Pickles, 2002). The multivariate Lagrange test suggested that model fit would be significantly improved only by including covariation paths between latent variables. Given that all measurement paths were significant at $p < .05$, in both robust and nonrobust estimates (Satorra & Bentler, 1990) and that residual estimates were quite small (i.e., RMSEA = .166), we conclude that the measurement model provided a reliable estimate of the constructs of interest (reliabilities were also high: Cronbach's $\alpha = .841$; Spearman's $\rho = .856$).

TABLE 9
Measurement and Structural Models Linking Lower Order Reading Processes to Reading Comprehension

Model		χ^2	*df*	*CFI*	*GFI*	*SRMR*	*RMSEA*	$\Delta\chi^2$	Δdf
1	Measurement model	1165.522	77	.737	.755	.296	.166	—	—
2	Structural model with direct paths linking fluency and accuracy to comprehension	239.336	70	.959	.938	.045	.069	928.186*	6
3	Structural model without direct paths linking Fluency and Accuracy to Comprehension	239.742	72	.960	.938	.046	.067	.406	2

Note. CFI = Comparative Fit Index; GFI = LISREL's Goodness of Fit Index; SRMR = Standardized Root Mean Squared Residual; RMSEA = Root Mean Squared Error of Approximation. The critical value for the χ^2 statistic in structural model comparisons 1 versus 2 was 12.59 (with 6 *df* at $p < .05$). The respective critical value for the comparison of Models 2 versus 3 was 5.99 (with 2 *df* at $p < .05$).

*Significant χ^2 statistic at $p < .05$

SRMR = .045). Because this model does not account for age (grade), the analysis was repeated with residualized estimates of each parameter, controlling for grade level, to provide an age-sensitive model. Results from the residualized model (see Figure 2) showed, again, that the data fit well the hypothesized relations, $\chi^2(70, N = 534) = 225.658$, $p < .001$ (CFI = .951, NNFI = .936, IFI = .951, SRMR = .046). There were no statistically significant differences in fit between the residualized and the non-residualized model.

The significant chi-square statistic for both models was not considered crucial to model fit as, in this case, the statistic was heavily influenced by excessive power (Cohen, 1992; Onwuegbuzie, Levin, & Leach, 2003) due to the large number of participants ($N = 534$). Thus, what constituted evidence regarding acceptable model fit was fit statistics above .90 and residual estimates less than .10 (Hu, Bentler, & Kano, 1992; Loehlin, 1987).

As shown in Table 9 (Model 2), fit statistics of the first structural model are well above the .90 cutoff point, suggesting acceptable model fit. However, the two structural paths linking accuracy and fluency to comprehension are miniscule, and the Wald test suggests that their withdrawal would not affect the magnitude of the chi-square statistic; that is, removing them would not decrease model fit. Thus, Model 3 tested the first structural model by dropping those two paths. Once again, model fit was acceptable, $\chi^2(72, N = 534) = 239.742$, $p < .001$ (CFI = .960, NNFI = .949, IFI = .960, SRMR = .046), and because the two structural models were not different from each other using chi-square difference tests, the simpler model is

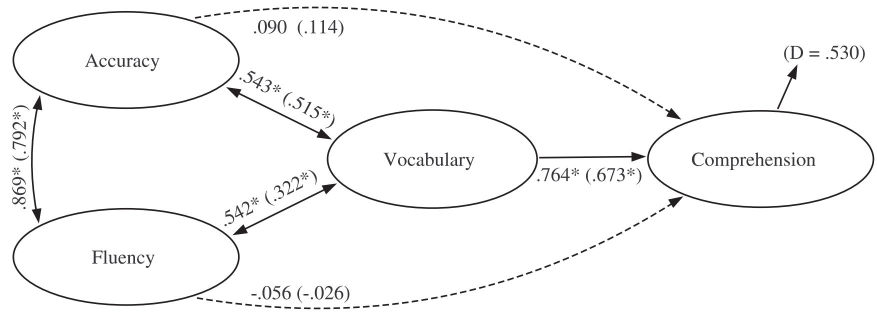

FIGURE 2 Structural equation model from the entire sample predicting reading comprehension from the simultaneous contribution of vocabulary, accuracy, and fluency. Unidirectional paths reflect standardized structural weights. Bidirectional paths indicate between-construct covariations. Model fit using robust estimates was, Satorra–Bentler, S–B $\chi^2(70, N = 534) = 209.910, p < .001$ (CFI = .960, NNFI = .948, IFI = .960, RMSEA = .062). The estimates in the parentheses represent effects after all variables were residualized for grade level. Thus, they represent a purer analysis of the effects. Using the residualized estimates, model fit was again acceptable, S–B $\chi^2(70, N = 534) = 200.885, p < .001$ (CFI = .949, NNFI = .934, IFI = .950, RMSEA = .060). The *D* value represents residual variance of the latent comprehension factor for the model that was residualized for age (* indicates paths that are significant at $p < .05$). Indicators are not shown for simplicity.

preferred for parsimony (following Kline, 1998). Thus, the final structural model explains variance in comprehension as a function of vocabulary only; the latter is defined by its covariation with fluency and accuracy.[3]

Invariance of slopes by grade level. Two sets of analyses were run to estimate the invariance of slopes across grade level. The first set of analyses tested the equivalence of the paths linking accuracy and fluency to comprehension. Although these paths were not significant for the aggregated sample, we tested the hypothesis that, for younger students, comprehension may be a function of accuracy or fluency (or both) and that those effects would likely diminish for older students. Thus, the equivalence of the paths between accuracy and comprehension and fluency and comprehension was tested across grade using latent variables. Results indicate that overall model fit is acceptable, $\chi^2(217, N = 534) = 402.075, p <$

[3]An alternative model was tested in which accuracy and fluency were linked to comprehension with direct paths, whereas vocabulary was linked to comprehension directly as well as indirectly via accuracy and fluency; that is, the order of vocabulary and word-level skills was reversed with respect to comprehension. This alternative model was found to be equivalent to Model 2. However, it was not preferred because (a) it lacked a parsimonious interpretation, and (b) comprehension was more strongly correlated with vocabulary than with accuracy and fluency; thus supporting a closer connection as long as structural models remain equivalent.

.001 (CFI = .943, NNFI = .928, IFI = .944, SRMR = .062), but that a set of equality constraints was significantly different across two grade groups. Specifically, there are significant differences between second graders and fourth graders regarding the path linking accuracy to comprehension. Second graders relied significantly more heavily on decoding accuracy to comprehend text (b = .334), whereas for fourth graders the respective path is significantly weaker (b = .171), suggesting that other processes were responsible for their comprehension, $\chi^2(1, N = 534) = 4.239$, $p < .05$.

The second set of analyses tested the hypothesis that the best fitting structural model (without the paths linking accuracy and fluency to comprehension) holds for all three age groups (i.e., Grades 2–4). The multi-group latent variable model specified constraints between (a) vocabulary and comprehension, (b) vocabulary and fluency (covariation), and (c) vocabulary and accuracy (covariation). The overall model fit the data well, $\chi^2(225, N = 534) = 415.523$, $p < .001$ (CFI = .941, NNFI = .929, IFI = .942, SRMR = .070). Figure 3 displays the coefficients for each grade group. Invariance of slopes was tested using a series of chi-square tests for all three group combinations (Grade 2 vs. Grade 3, Grade 2 vs. Grade 4, and Grade 3 vs. Grade 4). Results indicate that none of the structural paths were different from each other at any grade level using either the univariate or multivariate tests. Thus, the processes that contribute to reading comprehension seem to be invariant across grade. Specifically, it appears that reading comprehension is a function of the direct effect of vocabulary, and the latter is significantly related to fluency and accuracy.

Invariance of latent means by grade level. The fact that the slopes between groups are not different from each other indicates that the processes that explain the variability in comprehension function in approximately similar ways across grades. However, the three grade groups could be different from each other at the mean level of the latent variables. Thus, a multi-group latent means analysis was run in EQS to evaluate the hypothesis that the latent means (in accuracy, fluency, vocabulary, and comprehension) were invariant across grade. A significant advantage of this modeling is that means are compared at the latent level; that is, by the variables defined by multiple indicators (in our case, multiple tests). Three dummy variables were created to define group membership, with zero indicating the lower grade in each comparison (see Figure 4). Thus, the dummy variable for comparing Grades 2 and 3 had a value of 0 for Grade 2 and a value of 1 for Grade 3. A positive coefficient in any latent variable would indicate how much higher the mean at Grade 3 was compared to the mean of Grade 2. Fit indexes above .90, significant b-coefficients linking indicators to latent variables, and residual values below .05 were the criteria for determining acceptable model fit (Hu & Bentler, 1995; Hu & Bentler, 1998a, 1998b).

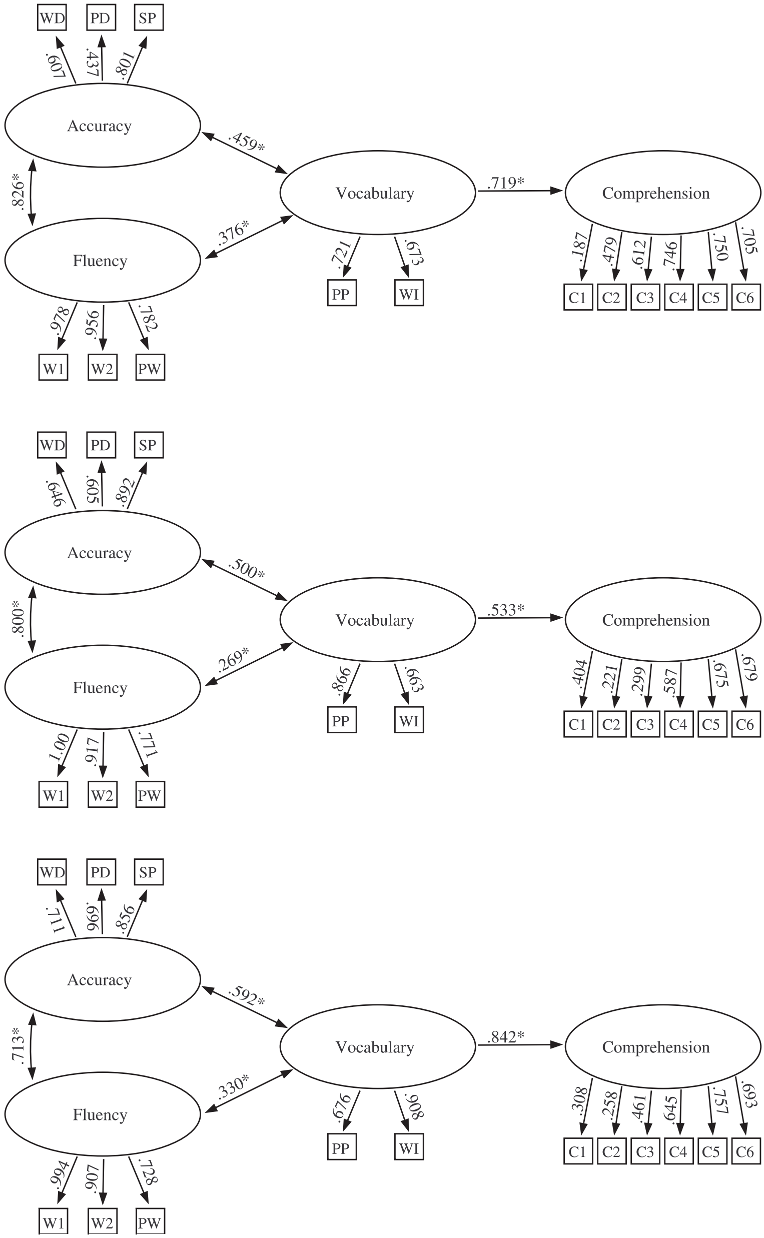

FIGURE 3 Structural model predicting reading comprehension from vocabulary, accuracy, and fluency. All structural paths are significant. Model fit as a function of the simultaneous estimation of all three grade groups is as follows, $\chi^2(225, N = 179/166/169) = 411.265$, $p < .001$ (CFI = .943, NNFI = .930, IFI = .944, SRMR = .068), which is acceptable. The upper panel shows the model applied to Grade 2, the middle panel to Grade 3, and the lower panel to Grade 4 participants.

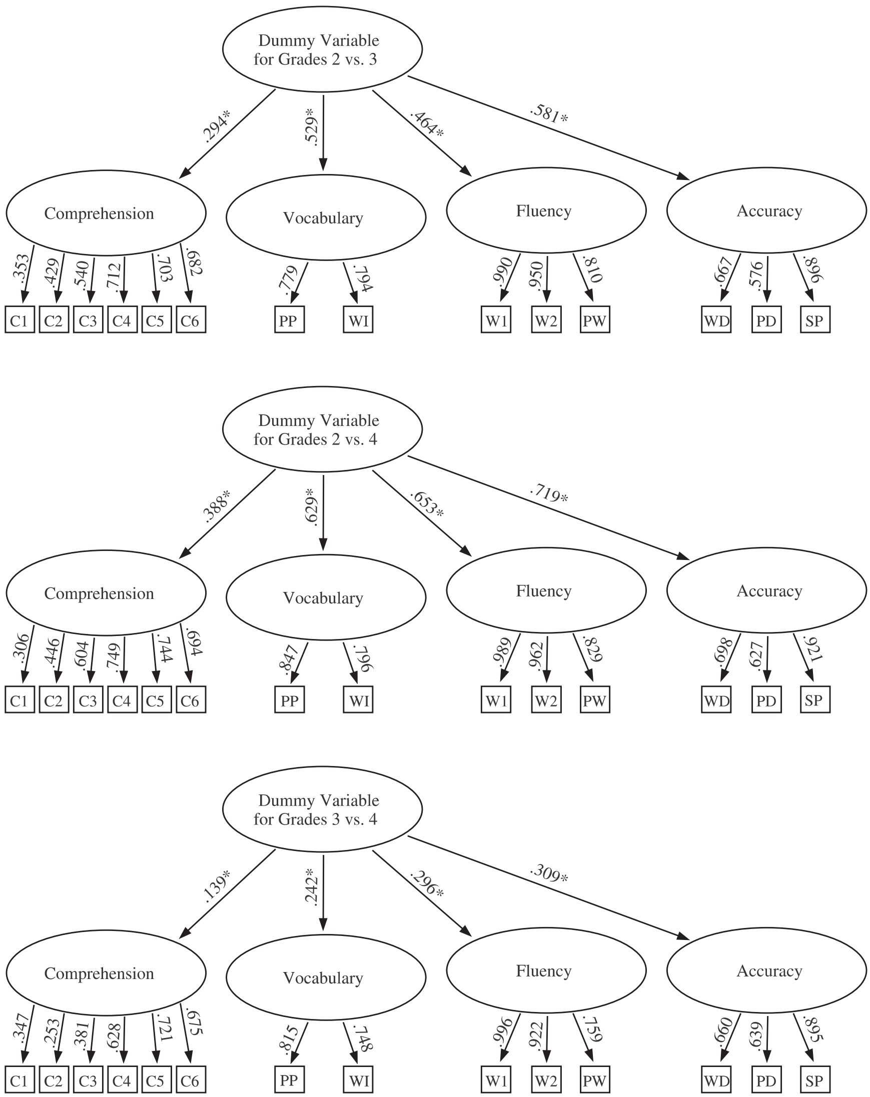

FIGURE 4 Latent means models across grade groups. Coefficients indicate differences in latent means as a function of grade comparison (* indicates paths that are significant at $p < .05$). The robust estimates were for Grade 2, Satorra–Bentler, S–B $\chi^2(81, N = 534) = 182.982$, $p < .001$ (CFI = .958, NNFI = .945, IFI = .958, RMSEA = .060); for Grade 3, S–B $\chi^2(81, N = 534) = 172.450$, $p < .001$ (CFI = .970, NNFI = .961, IFI = .970, RMSEA = .057); and for Grade 4, S–B $\chi^2(81, N = 534) = 188.397$, $p < .001$ (CFI = .944, NNFI = .927, IFI = .944, RMSEA = .063).

As shown in Figure 4 (top), there are statistically significant differences between Grade 2 and Grade 3 groups on all latent variables using robust or nonrobust solutions. Grade 3 students have significantly higher means in accuracy, fluency, vocabulary, and comprehension compared to Grade 2 students. Model fit is excellent, $\chi^2(81, N = 534) = 199.586$, $p < .001$ (CFI = .957, NNFI = .944, IFI = .957, SRMR = .045; Jaccard & Wan, 1992). All indicators defining the constructs are significant at $p < .01$ with loadings ranging between .353 and .99. With regard to the comparison between Grade 2 and 4 students (Figure 4, middle), significant differences between groups emerged across all constructs. Once again, the data fit the model very well, $\chi^2(81, N = 534) = 186.456$, $p < .001$ (CFI = .968, NNFI = .958, IFI = .968, SRMR = .052). All indicators defined appropriately their respective factors. Concerning the comparison between Grades 3 and 4, significant differences emerged across all latent variables (Figure 4, bottom) with the latter group having higher means, $\chi^2(81, N = 534) = 208.106$, $p < .001$ (CFI = .942, NNFI = .925, IFI = .943, SRMR = .049). Again all indicators defining the latent constructs are significant. These findings suggest that changes across the four latent variables are substantial for each grade. However, the processes seem to unfold in similar ways to predict reading comprehension (i.e., there was invariance of slopes but lack of invariance in intercepts).

Profiles of Poor Performance

The profiles of the poorest comprehenders, defined as the 10th percentile on our reading comprehension measure, indicate that these children differ from the full sample mean primarily on a vocabulary measure. To examine the profiles, we calculated normalized *z* scores for each measure, separately for each grade, and then tested the mean *z* score of each measure for this lowest 10% of the sample, via *t* test, against zero. Table 10 shows the mean *z* scores and significance levels of the associated *t* statistics for these comparisons in each grade. Notably, by Grade 4 the extremely poor comprehenders differ from the full sample to some extent in most measures (with the exception of RAN), including word-level skills such as reading and spelling accuracy, but the differences are greatest and most consistent for the vocabulary measures.

Conversely, the profiles of the poorest decoders, defined on the basis of inaccurate pseudoword reading (at the 10th percentile; also shown in Table 10), do not differ from the general population in the vocabulary or reading comprehension measures, but they do differ substantially and consistently in all word-level measures of accuracy and fluency. Finally, slow readers, defined on the basis of the word fluency measure (at the 10th percentile), also do not differ from the full sample in vocabulary or comprehension. Overall, slow readers present a very similar profile to that of the poorest decoders, with the exception that slow readers are also significantly slower in rapid naming, whereas poor decoders are not.

TABLE 10
Mean z Scores of Subgroups of Children Formed on the Basis of Poor Performance (10th Percentile) in Selected Measures

	Poor Comprehenders			Poor Decoders			Slow Read ers		
Measure	*Grade 2*[a]	*Grade 3*[b]	*Grade 4*[c]	*Grade 2*[d]	*Grade 3*[e]	*Grade 4*[f]	*Grade 2*[a]	*Grade 3*[b]	*Grade 4*[f]
Word fluency	–0.53	–0.49	–0.88*	–0.41	–0.46*	–0.92*	–1.59***	–1.76***	–1.86***
Pseudoword fluency	–0.55	–0.35	–0.68*	_0.77**	–0.69***	–1.03**	–1.16***	–1.13***	–1.27***
WISC blocks	–0.72**	–0.45	–0.70*	–0.33	–0.35	–0.51	.012	–0.23	–0.59
WISC vocabulary	–0.81***	–0.67	–1.06***	0.03	–0.11	–0.64	–0.31	–0.26	–0.39
RAN letters	–0.06	–0.08	–0.47	–0.17	–0.18	–0.53*	–1.04**	–1.24***	–0.89***
Word accuracy	–0.66*	–0.75	–0.94*	–0.60	–0.88**	–1.40***	–0.84**	–0.82*	–0.98
Pseudoword accuracy	–0.29	–0.23	–0.80*	–1.54***	–1.56***	–1.98***	–0.25	–0.67*	–0.95*
Reading comprehension	–1.73***	–2.10***	–2.09***	0.13	–0.10	–0.61	–0.64*	–0.55	–0.78
PPVT	–0.96**	–1.07*	–1.53***	0.18	–0.14	–0.87	–0.21	–0.18	–0.80
Spelling	–0.56	–0.91*	–1.06*	–0.20	–0.67**	–1.18**	–0.80**	–1.43***	–1.57***

Note. WISC = Wechster Intelligence Scales for Children; RAN = rapid automatized naming of letters; PPVT = Peabody Picture Vocabulary Test; Subgroups based on poor performance in reading comprehension (Poor Comprehenders), pseudoword accuracy (Poor Decoders), and word fluency (Slow Readers). Normalization to *z* scores, as well as calculation of percentiles, done separately for each grade. Group Ns are unequal due to ties at the 10th percentile.

[a]$N = 21$. [b]$N = 18$. [c]$N = 20$. [d]$N = 19$. [e]$N = 23$. [f]$N = 17$.

$p < .01$. $^{**}p < .001$. $^{***}p < .0001$. (by *t* test with respect to zero).

DISCUSSION

Our findings appear consistent with the view that the capacity to read with comprehension may largely depend on a dimension of lexical skill that is mutually reinforced and develops interactively with both word-level reading skill and with reading comprehension ability. Specifically, our measures show that, even though word reading skill remains modestly but significantly correlated with reading comprehension through the fourth grade (see Table 4), this correlation all but disappears once nonverbal ability and vocabulary are taken into account (see Tables 5 and 6). Therefore, most of the variance contributed to reading comprehension by decoding skill is not independent from lexical skills. Latent variable modeling strongly supports this conclusion, moderated only insofar as second graders still show a small, but significant, direct effect of decoding on comprehension.

Because lexical skill is assessed orally and not via reading (both the expressive and the receptive measure), there is little unique variance from the processing of printed words that affects the comprehension of text. This finding seems to run counter to the spirit of the Simple View, which posits a print-independent, verbal-skills component and a print-dependent, word-decoding component of reading comprehension. Instead, our findings are consistent with the view that a reciprocal relation exists between lexical skills and word-level decoding (and associated fluency) in supporting the comprehension of text (cf. Perfetti & Hart, 2002).[4]

The mediation, by vocabulary, of this reciprocal effect, is consistent with the finding of Seigneuric and Ehrlich (2005) that Grade 1 vocabulary contributes unique variance to Grade 3 reading comprehension, after taking into account age and the autoregressive effect. According to Seigneuric and Ehrlich, reading comprehension affects the growth of vocabulary knowledge. The reciprocal relation is consistent with the autoregressive effect of reading comprehension taking up all the variance from vocabulary (Goff et al., 2005).

In a similar vein, in considering the difficulties of less-skilled comprehenders, Oakhill et al. (2003) pointed to the richness of a child's semantic representations, rather than vocabulary per se, as being the crucial factor. They suggested that good comprehension may be fundamental to the ability to learn new vocabulary instead of the more common assumption that intelligence or vocabulary causes good comprehension. Whether vocabulary is considered an enabling factor or a consequence of good comprehension, either way the connection must be made via strong lexical representations of high quality. As for the interdependence among decoding and comprehension, Hagtvet (2003) suggested that it arises from "a more extensive

[4]The equivalence between a model in which lexical skills mediate the effect of decoding on comprehension, and an alternative model in which accuracy and fluency mediate the effect of vocabulary, further strengthens the interpretation in favor of a reciprocal interdependence among lexical and decoding skills.

impact from oral language to reading comprehension than what is mediated by decoding" (p. 524).

It should be clarified that our findings should not be taken to indicate that reading comprehension is determined solely by lexical skills. On the contrary, the importance of cognitive processes related to memory, monitoring, integration, and other verbal skills has been well established (e.g., Cain et al., 2004; Oakhill et al., 2003). The aim of our study was not to evaluate either the Simple View or the lexical quality hypothesis. Rather, our goal was to examine the common variance shared by decoding and comprehension, hoping to shed light on their relation with lexical skills as indexed by vocabulary measures. In particular, we do not use vocabulary as a proxy for verbal skills to replace listening comprehension in the Simple View. We focus entirely on the contribution from the print-dependent component of the model, without examining the linguistic comprehension component, and we observe that what decoding would contribute to comprehension is largely taken up by variance from non-decoding measures; namely, vocabulary. It should be emphasized that decoding is not claimed to be unrelated to comprehension, but that decoding variance is shared with vocabulary variance in accounting for differences in comprehension. Because vocabulary is more strongly correlated with comprehension than is decoding, we may thus say that vocabulary dominates decoding in accounting for variance in reading comprehension. The results of the latent variable model analysis indicate that one can justifiably claim that vocabulary mediates decoding effects on comprehension.

It is important to point out that neither vocabulary measure alone (PPVT or WISC) could account for nearly as much variance in comprehension as the two combined. In every analysis in which a vocabulary measure was found of highest importance (e.g., the regressions of comprehension), the other vocabulary measure was typically second highest. Thus, a valid "lexical" construct needs at least two complementary indicators, and it cannot be precluded that our measures of receptive and expressive vocabulary may be complemented, or even replaced, by additional or more appropriate (or both) indexes of lexical representation quality.

The particular vocabulary measures used in this study are not directly relevant for comprehension, insofar as the content of the texts to be read does not include very low frequency or otherwise difficult words. Therefore, our vocabulary measures, taken as indexes of the size of the child's spoken vocabulary, do not appear to be crucial for scoring high on the comprehension task. Given that none of the words included in the word reading tasks appeared in the vocabulary tasks, it seems reasonable to suggest that the "vocabulary" construct that mediates the effect of word-level reading on reading comprehension is, in fact, an index of the quality of lexical representations, which may underlie both lexical access from the printed words as well as the construction of the meaning of text, by affecting the ability to recognize printed words with a high degree of accuracy and automaticity (cf. Dixon, LeFevre, & Twilley, 1988; where vocabulary contributed large amounts of variance in reading comprehension, reading rate, and reading span).

To the extent that a rich combined expressive and receptive vocabulary, as measured by high performance on our vocabulary tests, is acquired and maintained through practice in verbal skills and in particular with written language, primarily via reading, then vocabulary measures will be strongly positively correlated with strong connections between orthographic, phonological, and semantic specifications in the lexical networks; hence with high lexical quality. Because of this contingency, vocabulary will also be strongly correlated with word reading skills and, in particular, with the contribution of word reading skills to reading comprehension. Therefore, whether support is found for the Simple View or for the lexical quality hypothesis may be expected to be largely a function of what skills are examined. From an interactive processing point of view, strengthening of common pathways ensures that common patterns will emerge in measures assessing different aspects of common processes.

Thus, experience with print is hypothesized to affect both the lexical representations themselves and the processing of orthographic forms that permit access to the lexical representations. If the processing of orthography is sufficiently consistent (as it is in Greek) so that lexical representations are the most demanding of these two components, then it is possible that lexical quality will be the factor limiting comprehension most, even when orthographic processing itself is inadequate as well. Therefore, although lack of experience with print (or inappropriate or inadequate processing) may deprive children from a normally valuable aid toward developing fluent decoding and word recognition, reduced experience with print may also deprive them from a useful means to improve the quality and strength of lexical representations. If orthographic processing is inadequate or incorrect, then the resulting orthographic specifications will suffer, as will their connections with the corresponding phonological and semantic representation in the lexicon. Thus, a processing limitation (in the early stages of learning to read) turns into a resource limitation (in poor lexical representations) that limits reading comprehension.

In addition to a partial separability of comprehension from word-level accuracy and fluency at the low end of each skill, the profiles of poor performance in our sample indicate that Matthew effects may be present. One clue comes from the selection variable for the examination of the corresponding profile (reading comprehension for the “poor comprehenders,” pseudoword accuracy for the “poor decoders,” and word fluency for the “slow readers”). This measure is expected to show the largest difference from the sample mean due to regression to the mean for the other measures. However, this difference is not necessarily expected to increase at higher grades, as is apparent in Table 10. This means that the poorest comprehenders (or decoders or slowest readers) are farther away from the mean for their age group at higher grades.

Moreover, the difference of the other measures, in each profile, from the sample mean, is also not necessarily expected to increase at higher grades, but it does seem to do so. For example, as Table 10 shows, measures tapping important skills, such as word fluency and spelling, are increasingly poorer for poor comprehenders and eventually attain statistical significance in Grade 3 or 4. If this pattern can be systematically confirmed in future studies, it means that less skilled readers fall behind in several related domains as grades progress, as would be predicted from an interactive view of lexical, verbal, and reading skills.[5] We do not pursue these differences further in this study, because the cross-sectional sample does not permit clear demarcation of developmental effects.

Among European languages, English appears to be an exception, in that the developmental trajectory of reading skills is at least severely delayed, if not qualitatively distinct, compared to the other languages. This difference seems to a large extent attributable to orthographic inconsistency, although the effects of educational practices remain to be fully explored (Seymour, Aro, & Erskine, 2003; see also Spencer, 2000). Therefore it must be determined empirically if these findings for Greek, strong already by the fourth grade, can be extended to other languages with consistent orthography. It should also be examined whether they can be applied to English, perhaps with a lag accounting for the overall delay in learning to read English as compared to languages with more consistent orthography. One possible prediction[6] is that in English, in which semantic knowledge is necessary to support the identification of irregular words, one would expect to find an even stronger relation between word identification and vocabulary than in Greek, with the relatively transparent orthographic system, in which sight vocabulary is strongly supported by graphophonemic consistency.

The mediation of lexical representations in the relation between word-level reading and text comprehension suggests a possible explanation for the paradox introduced by the empirical dissociation, increasingly at higher grades, between decoding and comprehension. Specifically, the question is, how can a child who decodes very poorly and reads a text slowly, effortfully, and perhaps with many errors, understand this text with little or no difficulty when not pressed to do it quickly? There is evidence that this (dyslexic) child uses context to support comprehension much more than a skilled decoder would (Nation & Snowling, 1998). The use of context means that semantic networks are brought to bear strongly on the construction of meaning; not because lexical activation is strongly facilitated

---

[5]Interestingly, if decoding ceases to have a direct effect on comprehension and yet poor comprehenders fall behind in decoding, among other skills, due to limited reading practice and reciprocal interactions, one may suggest that the opposite of the Simple View is the case: Instead of poor decoding preventing adequate comprehension, it is poor comprehension that holds back development of fluent decoding.

[6]We thank an anonymous reviewer for pointing this out.

from the bottom up, but because the semantic network itself is strong. Such compensatory use of context may hinge, at least in part, on additional verbal and perhaps nonverbal cognitive skills; hence poorly decoding good comprehenders typically perform well in measures of memory, as well as verbal and nonverbal intelligence (Lesaux, Pearson, & Siegel, 2006). Conversely, a poor comprehender with strong decoding skills may fail to construct text meaning adequately, despite fluent orthographic processing. This may be because the semantic network is somehow insufficiently cohesive or effective, and this will be systematically reflected in vocabulary measures but also often in other verbal and perhaps nonverbal cognitive measures (Cain et al., 2000; Oakhill et al., 2003).

If this view is on the right track, there are implications for the assessment of comprehension, because educationally meaningful assessments may have to go beyond the single paragraph with multiple-choice questions at the end. Such minimal testing does indeed give an indication of how well a child can derive some meaning from the text, but it falls short of informing us how completely and efficiently this meaning is derived and how it can be put to use in a more demanding circumstance—for example, in combination with other knowledge, in reformulation and retelling of the story, or in successfully following specific instructions. In agreement with Fletcher (2006), we suggest that multidimensional approaches to reading comprehension assessment are becoming increasingly necessary, in psychoeducational settings as well as in cognitive research.

An important caveat concerns the nature of the reading comprehension measure employed, as it has now been repeatedly demonstrated that different instruments may probe somewhat different sets of skill domains and present distinct profiles of interrelations with other cognitive, reading, and language measures. Cutting and Scarborough (2006) found substantial differences in the amounts of variance in comprehension attributed to oral language and to decoding among three widely used standardized tests of reading comprehension, even though all three employed a passage-plus-questions format and differed in relatively minor ways. Despite these differences, and in agreement with our findings, Cutting and Scarborough found a "very substantial amount of shared variance between word recognition/decoding and oral language measures when comprehension scores are predicted" and suggested that "their largely combined, rather than unique, contribution is not entirely clear and merits further investigation" (p. 293).

Differences between a cloze-format test and two passage-plus-questions tests of reading comprehension were found by Francis, Fletcher, Catts, and Tomblin (2005) in the amount of variance attributed to decoding versus oral language measures. The cloze procedure more strongly correlated with decoding, whereas the passage–question tests were more strongly related to oral language measures (see also Francis et al., 2006). Similarly, Spear-Swerling (2004) found word accuracy more strongly related to a cloze test than to a passage-plus-questions test, and Hagtvet (2003) observed that vocabulary was most strongly related to their story-retelling measure of reading comprehension, whereas phonemic awareness and syntax were most strongly related to their cloze task. In the sentence verification tasks of Diakidoy, Stylianou, Karefillidou, and Papageorgiou (2005), reading and listening comprehension became more strongly correlated after Grade 2 for narrative but not for expository texts.

In addition to such statistical and conceptual considerations affecting choice of a reading comprehension test (Cross & Paris, 1987) and in light of new instruments appearing specifically to assess components of reading comprehension (e.g., Hannon & Daneman, 2001; Francis et al., 2006), it is important to examine more closely the cognitive processes underlying performance in each test, to present in detail the procedures and properties of each test used, and to issue a caveat limiting generalization of conclusions to the types of tests employed.

In conclusion, we suggest that an apparent conflict may arise between the Simple View and the lexical quality hypothesis if component processes are seen as clearly distinct and independent from one another. This artificial separation may lead to empirical associations and dissociations among variables that depend more on which set of measures were taken in a given study, and in what order they were considered in a regression model, than on their actual role in the cognitive processes of reading. In agreement with Nation and Snowling (2004) and Braze et al. (in press), we propose that emphasizing the interrelation of processing components, from a connectionist modeling approach, may offer a fruitful direction for understanding the reading process, including aspects now emphasized either by the Simple View or by the lexical quality hypothesis. Future studies should administer wide ranging assessment batteries, longitudinally, to capture all reliable variance related to reading efficiency, and employ structural developmental modeling to test alternative proposals to partition the variance and assign causality while emphasizing the interactive and reciprocal character of component skill development.

ACKNOWLEDGMENTS

Partial preliminary results of the studies reported here have been reported in July 2005 at the International School Psychology Colloquium, in Athens, Greece, and in December 2005 at the 10th Panhellenic Conference of Psychological Research, in Ioannina, Greece. We are grateful to Dave Braze for comments on this article.

REFERENCES

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19,* 933–958.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34*(6), 1373–1399.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*(4), 506–521.

Bentler, P. M. (2004). *EQS structural equations program manual.* Los Angeles: BMDP Statistical Software Inc.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin, 130*(6), 858–886.

Braze, D., Tabor, W., Shankweiler, D., & Mencl, E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabil*ities*, 40*(3), 226–243.

Cain, K., Oakhill, J., & Bryant, P. (2000). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing, 13,* 31–56.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96,* 31–42.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Cross, D. R., & Paris, S. G. (1987). Assessment of reading comprehension: Matching test purposes and test properties. *Educational Psychologist, 22,* 313–332.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10,* 277–299.

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

Denckla, M. B., & Rudel, R. G. (1974). "Rapid automatized naming" of pictured objects, colors, letters, and numbers by normal children. *Cortex, 10,* 186–202.

Diakidoy, I.-A. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology, 26,* 55–80.

Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80*(4), 465–472.

Dunn, G., Everitt, B., & Pickles, A. (2002). *Modeling covariances and latent variables using EQS.* New York: Chapman & Hall.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test–Revised.* Circle Pines, MN: American Guidance Service.

Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10,* 323–330.

Francis, D. J., Fletcher, J. M., Catts, H. W., & Tomblin, J. B. (2005). Dimensions affecting the assessment of reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369–394). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading, 10,* 301–322.

Georgas, D. D., Paraskevopoulos, I. N., Bezevegis, I. G., & Giannitsas, N. D. (1997). *Elliniko WISC–III: Wechsler klimakes noimosinis gia pedia* [Greek WISC–III: Wechsler intelligence scales for children]. Athens, Greece: Ellinika Grammata.

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing, 18,* 583–616.

Gorsuch, T. (1983). *Factor analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10.

Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing: An Interdisciplinary Journal, 16,* 505–539.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93,* 103–128.

Hatzigeorgiu, N., Gavrilidou, M., Piperidis, S., Carayannis, G., Papakostopoulou, A., Spiliotopoulou, A., et al. (2000). Design and implementation of the online ILSP corpus. *Proceedings of the second international conference of language resources and evaluation (LREC)* (Vol. 3, pp. 1737–1740). Athens, Greece.

Holopainen, L., Ahoren, T., & Lyytinen, H. (2001). Predicting delay in reading achievement in a highly transparent language. *Journal of Learning Disabilities, 34*(5), 401–413.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127–160.

Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Hu, L. T., & Bentler, P. M. (1998a). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 55–72.

Hu, L. T., & Bentler, P. M. (1998b). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Quantitative Methods in Psychology, 112,* 351–362.

Jaccard, J., & Wan, C. K. (1992). *LISREL approaches to interaction effects in multiple regression.* Thousand Oaks, CA: Sage.

Jiménez González, J. E., & Hernández Valle, I. (2000). Word identification and reading disorders in the Spanish language. *Journal of Learning Disabilities, 33*(1), 44–60.

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing, 19,* 339–361.

Joshi, R. M. (2005). Vocabulary: A critical component of comprehension. *Reading & Writing Quartertly, 21,* 209–219.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21,* 85–97.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Landerl, K. (2001). Word recognition deficits in German: More evidence from a representative sample. *Dyslexia, 7,* 183–196.

Landi, N. (2005). *Behavioral and electrophysiological investigations of semantic processing in skilled and less-skilled comprehenders.* Unpublished doctoral dissertation, University of Pittsburgh, PA.

Lesaux, N. K., Pearson, M. R., & Siegel, L. S. (2006). The effects of timed and untimed testing conditions on the reading comprehension performance of adults with reading disabilities. *Reading and Writing, 19,* 21–48.

Loehlin, J. C. (1987). *Latent variable models: An introduction to factor, path, and structural analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lyon, G. R., Fletcher, J. M., & Barnes, M. C. (2002). Learning disabilities. In E. J. Mash & R. Barkley (Eds.), *Child psychopathology* (2nd ed., pp. 520–586). New York: Guilford.

Megherbi, H., & Ehrlich, M.-F. (2005). Language impairment in less skilled comprehenders: The on-line processing of anaphoric pronouns in a listening situation. *Reading and Writing, 18,* 715–753.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665–681.

Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development, 69*(4), 996–1011.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27,* 342–356.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18,* 443–468.

Onwuegbuzie, A., Levin, J. R., & Leach, N. L. (2003). Do effect sizes measure up? A brief assessment. *Learning Disabilities: A Contemporary Journal, 1,* 37–40.

Padeliadu, S., & Sideridis, G. D. (2000). Discriminant validation of the Test of Reading Performance (TORP) for identifying children at risk of reading difficulties. *European Journal of Psychological Assessment, 16,* 139–146.

Perfetti, C. A., & Hart, L. (2001). The lexical bases of comprehension skill. In D. S. Gorfien (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67–86). Washington, DC: American Psychological Association.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam/Philadelphia: John Benjamins.

Porpodas, C. D. (1999). Patterns of phonological and memory processing in beginning readers and spellers of Greek. *Journal of Learning Disabilities, 32,* 406–416.

Protopapas, A., & Skaloumbakas, C. (2007). Traditional and computer-based screening and diagnosis of reading disabilities in Greek. *Journal of Learning Disabilities, 40,* 15–36.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* Santa Monica, CA: RAND.

Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis, 10,* 235–249.

Sattler, J. M. (1982). *Assessment of children's intelligence and special abilities* (2nd ed.). Boston: Allyn & Bacon.

Savage, R. (2006). Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. *Scientific Studies of Reading, 10,* 143–164.

Seigneuric, A., & Ehrlich, M.-F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing, 18,* 617–656.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94,* 143–174.

Sideridis, G. D., & Padeliadu, S. (2000). An examination of the psychometric properties of the Test of Reading Performance using structural equation modeling. *Psychological Reports, 86,* 789–801.

Spear-Swerling, L. (2004). Fourth graders' performance on a state-mandated assessment involving two different measures of reading comprehension. *Reading Psychology, 25,* 121–148.

Spencer, K. (2000). Is English a dyslexic language? *Dyslexia, 6,* 152–162.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21,* 590–604.

Tressoldi, P. E., Stella, G., & Faggella, M. (2001). The development of reading speed in Italians with dyslexia: A longitudinal study. *Journal of Learning Disabilities, 34*(5), 414–417.

Tunmer, W. E., & Hoover, W. A. (1992). Cognitive and linguistic factors in learning to read. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 175–214). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*(3), 4–12.