Read Writ
DOI 10.1007/s11145-011-9336-5

The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school

C. Patrick Proctor · Rebecca D. Silverman · Jeffrey R. Harring · Christine Montecillo

© Springer Science+Business Media B.V. 2011

Abstract The present study investigated the role of vocabulary depth in reading comprehension among a diverse sample of monolingual and bilingual children in grades 2–4. Vocabulary depth was defined as including morphological awareness, awareness of semantic relations, and syntactic awareness. Two hundred ninety-four children from 3 schools in a Mid-Atlantic district and 3 schools in a Northeastern school district participated in the study and were assessed at the beginning and end of one school year on a wide variety of language and literacy measures. Bilingual children were assessed in English and Spanish. A latent difference score model that assessed change in a latent indicator of English reading comprehension from Time 1 (Fall) to Time 2 (Spring) was tested with results showing that vocabulary depth measures made significant contributions to initial status, but not change, in reading comprehension over and above between-subjects factors (grade, ethnicity, language status) and baseline control within-subject factors (word identification and vocabulary breadth). There was no added contribution of Spanish language measures to English reading comprehension among the bilingual students.

Keywords Vocabulary · Depth · Bilingual · Spanish · Reading comprehension

C. P. Proctor (✉) · C. Montecillo
Department of Curriculum & Instruction, Lynch School of Education, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467, USA
e-mail: proctoch@bc.edu

R. D. Silverman
Department of Special Education, University of Maryland, College Park, MD, USA

J. R. Harring
Department of Measurement, Statistics, & Evaluation, University of Maryland, College Park, MD, USA

Published online: 12 August 2011

For all students, mastering basic word-level skills is a necessary first outcome toward the ultimate goal of reading for comprehension and meaning. Comprehension is especially difficult in the upper elementary and secondary grades due to the increasing linguistic complexity of texts. At notable risk for comprehension difficulties in the US are Latina/o bilingual children who grow up in homes where Spanish is spoken, as evidenced by achievement discrepancies with their Anglo counterparts on standardized assessments (NCES, 2007) and school dropout rates (Child Trends Data Bank, 2005). Of particular relevance for the current study, well-developed vocabulary knowledge has been established as a fundamental component of the linguistic proficiency necessary for facile reading comprehension among both monolingual and bilingual learners from a variety of language backgrounds, including Spanish (Geva, 2006; Lesaux, Rupp, & Siegel, 2007; Hoover & Gough, 1990; Proctor, Carlo, August, & Snow, 2005; Snow, Burns, & Griffin, 1998). However, most vocabulary research has focused on the effects of *breadth* of vocabulary (i.e., the number of known words) on reading outcomes. The goal of the current study was to explore how *depth* of vocabulary (i.e., the richness of word understandings) contributes to reading comprehension among a sample of second, third, and fourth grade English monolingual and Spanish–English bilingual students, and whether, among the bilingual students, Spanish proficiency mediated those relationships.

What is depth of vocabulary knowledge?

The recent attention of researchers to vocabulary development and instruction has been devoted less to depth of vocabulary and more to breadth of vocabulary. It is indeed astonishing to consider the sheer quantities of words that students need to know, on average, for twelfth grade preparedness (approximately 40,000–50,000; Graves, 2006; Nagy & Herman, 1985). It is also sobering to consider that, by the twelfth grade, those children who began school with limited vocabulary knowledge may know half as many academically relevant words as children who came to school with just average vocabulary knowledge (Biemiller & Slonim, 2001; Hart & Risely, 1995; White, Graves, & Slater, 1990). This sheer vocabulary size has been associated with reading and achievement for nearly a century (Pearson, Hiebert, & Kamil, 2007), and most reading comprehension studies operationalize vocabulary knowledge with vocabulary breadth measures. Measures that assess breadth of word knowledge seek out a single representation for a given item and are not sensitive to how context affects word meanings. On a breadth of vocabulary measure, for example, the word *table* would only be allowed a single meaning for a single context, which tells us relatively little about how deep a child understands the word’s meaning.

A student who has depth of word knowledge for the word *table*, however, is aware that one sits at a table, but in reading a science text or conducting an experiment, one also creates, reads, and/or interprets a table. Further, while *table* is a noun, its morphological derivation, *tabulate*, is a verb, and is thus used in a grammatically and syntactically different context. Thus, in order to create a table, one must tabulate relevant data, perhaps while seated at the dinner table. Far less attention has been paid to the investigation of this type of vocabulary depth, its contribution to reading comprehension, and how that relationship is relevant for instruction. Arguably, depth of word knowledge is a form of metalinguistic awareness, the effects of which have been established within and across languages for both reading and cognition (Kuo & Anderson, 2010; Bialystok, 2006; Hakuta, 1986; Hakuta & Diaz, 1985). It is this depth of word knowledge that is theorized to predict reading performance in the current research. While there are surely myriad ways in which to operationalize depth of vocabulary knowledge (see Perfetti, 2007; Perfetti & Hart, 2002 for a similar perspective, the *Lexical Quality Hypothesis*, as it pertains to adult readers), a review of the extant literature suggests that there are three related but distinct linguistic domains that affect the quality of students' word knowledge and reading comprehension: Morphology, semantics, and syntax. These three constructs of vocabulary depth are discussed in the following sections.

Morphology

Knowledge of morphology enables children to generalize the meaning of root words to their morphological derivations and apply them across syntactic contexts. Within the still limited research on depth of vocabulary, the domain of morphological awareness has received some preferred attention. Kuo and Anderson (2006) traced a morphological trajectory as moving from inflectional morphology in very young children (e.g., dog + s = dogs), to compound morphology (e.g., tooth + brush = toothbrush), and finally to derivational morphology (e.g., electric + ity = electricity), which is hypothesized to be the most advanced and reading-related stage of morphological awareness. It has been hypothesized that the relationship between morphological awareness and reading comprehension changes over time and becomes increasingly associated with reading comprehension (Anglin, 1993; Berninger, Abbott, Nagy, & Carlisle, 2009; Carlisle, 2007; Kieffer & Lesaux, 2007, 2008; Nagy, Berninger, & Abbott, 2006).

In their study of 607 4th through 9th grade monolingual English-speaking students, Nagy, Berninger, and Abbott (2006) found that, controlling for breadth of vocabulary in reading, morphological awareness exerted a small but significant effect on reading comprehension. Kieffer and Lesaux (2008) followed 87 Spanish–English bilingual fourth graders through their fifth grade year and assessed the independent contribution of morphological awareness on two measures of English reading comprehension (Woodcock Passage Comprehension and the Gates-MacGinitie). Controlling for a baseline model of reading that included breadth of vocabulary knowledge, Kieffer and Lesaux (2008) found that morphological awareness in fourth grade made independent contributions to both reading measures at fifth grade, explaining 7.8 and 6.1% additional variance, respectively.

Semantics

Knowledge of how words conceptually relate to one another characterizes semantics. Well-developed semantic understandings allow children to be sensitive to issues of polysemy (multiple meanings of words) across academic and social contexts (e.g., the multiple meanings of a seemingly simple word like *table*), how different words are linked by context (teacher/school), and child-specific understandings about words and concepts (i.e., background knowledge and predictions of word meanings). Theoretically, having greater semantic awareness facilitates comprehension by making it easier to determine the meaning of unknown words through context and to understand how related words convey meaning. And like morphology, the relationship between semantic awareness and reading comprehension appears to be developmentally bound.

In a study of second grade Dutch-Arabic bilinguals in Holland, Vermeer (2001) found correlations between semantic awareness and vocabulary breadth in Dutch to be so high that the two measures were indistinguishable. Tannenbaum, Torgesen, and Wagner (2006) found similar results in their work with 203 third grade students from predominantly English language backgrounds. While they were able to extract a distinct factor for semantic awareness, breadth of English vocabulary predicted the majority of variation in reading comprehension and correlated at .87 ($p < .001$) with semantics, such that only vocabulary breadth was predictive of reading.

Research with older students, however, paints a different picture. Nation and Snowling (2004) found that vocabulary breadth and semantics made significant contributions to reading comprehension for 13-year-old students. However, the authors did not test a single model that included both vocabulary breadth and semantics to determine whether semantic depth explained any variance beyond that of vocabulary breadth. Ouellette (2006), however, did conduct such an analysis and showed that among older students, semantic depth played a significant role in explaining reading comprehension, on control for vocabulary breadth and decoding skills. Proctor, Uccelli, Dalton, and Snow (2009) worked with a small sample of 36 monolingual and bilingual fifth grade students. They found that semantic depth was predictive of reading comprehension, on control for oral language proficiency (i.e., breadth of vocabulary and listening comprehension). Interestingly, the effect of semantics on reading was absent for students with more limited English oral skills (which included many bilingual students), moderate for students with average oral language proficiency, and strong for those students whose linguistic skills were in the 90th percentile.

Syntax

Knowledge of the structure of language is important for children to appropriately develop degrees of vocabulary depth. As children broaden their knowledge of particular words they learn the syntactic constructions in which these words typically and appropriately appear. Despite breadth of word knowledge, children who do not have well-developed syntactic awareness or who are bilingual and experience syntactic interference from heritage language knowledge may have difficulty developing robust word knowledge and comprehension (Nation, Clarke, Marshall, & Durand, 2004; Wolter, 2006).

In a study conducted with a mostly Anglo population of fifth grade students, Mokhtari and Thompson (2006) found that syntactic awareness was significantly related to fluency and comprehension, which indicated that lower levels of syntactic awareness were associated with decreasing fluency and comprehension. However, Cain (2007), in a study of 8–10 year old children in England, found that the influence of syntax on reading comprehension was mediated by other language and literacy skills, suggesting syntax had an indirect but not a direct influence on reading comprehension. Unfortunately, Cain's sample explicitly excluded students who did not speak English as their first language.

Low and Siegel (2005) studied the contribution of working memory, phonological processing, and syntactic awareness to the reading comprehension of 884 native English and 284 English-as-a-second-language (ESL) sixth-graders in Canada. They found that syntactic awareness added to the explanation of reading comprehension for both native English and ESL students, and that native English speakers outperformed ESL children in syntactic awareness. In contrast, there were no significant differences between these two groups on phonological processing and verbal working memory tasks, which also predicted comprehension. Considering these monolingual/bilingual differences, the relationship between syntax and comprehension warrants further investigation.

Spanish–English bilingualism and English reading

Recently, a small number of studies have sought to determine the degree to which Spanish language proficiency among Spanish–English bilingual and biliterate children predicts English comprehension outcomes. This research base draws on the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), which holds that word reading (decoding) and linguistic comprehension (oral language proficiency) interact with one another and predict the vast majority of variation in reading comprehension outcomes. Hoover and Gough (1990) tested this theory with a group of Spanish–English bilingual children and found strong support for this view, with decoding skill predicting larger shares of variation in comprehension among younger learners and oral language proficiency predicting the lion's share of variation among older students' reading comprehension. Hoover and Gough (1990), however, did not assess the students' Spanish language proficiency as a potential contributor to English comprehension.

Linguistic comprehension is often assessed with vocabulary breadth measures, which have tended to correlate negatively between Spanish and English (see, e.g., Carlisle, Beeman, Davis, & Spharim, 1999; Ordoñez, Carlo, Snow, & McLaughlin, 2002; Proctor, August, Carlo, & Snow, 2006). However, Proctor et al. (2006) worked with a sample of 135 Spanish–English bilingual fourth grade students and tested whether, on control for English predictors (pseudo-word reading, real word reading rate/accuracy, vocabulary knowledge, and listening comprehension), parallel Spanish variables predicted English reading comprehension. The authors found that Spanish vocabulary breadth exerted a significant main effect on English reading comprehension. Additionally, students who read in English with greater speed and accuracy (i.e., fluency) were those students whose English reading comprehension was most affected by positive changes in Spanish vocabulary breadth.

In related longitudinal research, Nakamoto et al. (2008) worked with 282 Spanish–English bilinguals from third through sixth grade, but found no main effect of Spanish vocabulary breadth on English reading. However, the authors did find that English reading comprehension was improved via the interaction of English decoding and Spanish vocabulary breadth, consistent with Proctor et al.'s (2006) findings. Finally, Mancilla-Martinez and Lesaux (2010) worked with 173 Spanish–English bilinguals and assessed whether performance on measures of word reading accuracy and productive vocabulary, administered annually in Spanish and English beginning at age 4.5, predicted English reading outcomes at age 11. The authors found no evidence to suggest that initial status and growth in Spanish language (vocabulary) or literacy (word reading) skills predicted variance in English reading comprehension at age 11. It is important to note that findings in this area, when present, have been quite modest, with approximately 2% of additional variance in English reading outcomes being explained by Spanish vocabulary breadth and its interactions with English text-level skills. Even so, given the conflicting findings in the literature, the present study ventured to add to the conversation in this arena.

The present study

In the present study, we investigated the role of vocabulary depth in reading comprehension for children in grades two through four. To our knowledge, there exist no studies of vocabulary depth that have sought to determine the relative contributions of three major indicators of depth of vocabulary knowledge (i.e., morphology, semantics, and syntax). Virtually all research in the realm of vocabulary depth has focused exclusively on one of these domains. We further wished to assess whether indicators of Spanish linguistic proficiency were associated with English comprehension once English language variables were added to a predictive model.

Our preliminary analysis included establishing a latent construct of reading comprehension and investigating grade, ethnicity, and language status effects on reading comprehension and other language and literacy variables. The sample included Anglo, African-American, and "other" monolinguals, as well as Spanish–English bilinguals. For the purposes of the current study, and per state educational regulations at both research sites, Spanish–English bilingual children who were classified as limited in English proficiency (LEP) by their schools were considered English language learners (ELLs) while those Spanish–English bilingual children who were no longer or never labeled as LEP were considered non-ELL. We asked three major research questions:

1. What is the nature of beginning levels of reading comprehension and its change over one academic year? What are the effects of grade and ethnolinguistic status?

2. What is the relationship between vocabulary depth (i.e., morphology, semantics, and syntax) and initial status and change in reading comprehension across an academic year, controlling for decoding, vocabulary breadth, grade, ethnicity, and language status, in grades 2–4?
3. Among the Spanish–English bilinguals, does Spanish language proficiency contribute to initial status and change in English reading comprehension across an academic year in grades 2–4?

Method

Participants

Two hundred ninety-four children participated in the study. There were 104, 95, and 95 participants in grades 2, 3, and 4, respectively. Fifty-six percent of the sample was comprised of monolingual English speakers. Forty-four percent of the students were Spanish–English bilinguals. Of the Spanish–English bilingual children, half (50%) were designated LEP (and thus ELL) by their schools. Students were recruited from one Northeastern site (n = 89) and one Mid-Atlantic site (n = 205) from one of three schools per site (n = 6 schools). The racial composition of the sample was 44% Latino, 35% African-American, 17% Anglo, and 4% from other ethnic backgrounds (all of whom were English monolinguals).

Further disaggregating the data revealed some interesting site differences that had implications for language status analyses. At the Northeastern site, there were 29 students in grades 2 and 3, and 31 students in grade 4. The bilingual non-ELL group comprised 31% of the students while 12% of the sample was ELL. The racial composition (43% Latino, 42% Anglo, 7% African American, and 8% other) indicated that the bilingual group was comprised entirely of Latino students while the monolingual group was predominantly Anglo.

At the Mid-Atlantic site, there were 77, 65, and 63 students in grades 2, 3, and 4, respectively. Comparable to the Northeastern site, 45% of the children were Spanish–English bilinguals. In contrast, however, a much higher 29% of the sample was bilingual ELL while 16% were bilingual non-ELL. The racial composition (45% Latino, 47% African American, 6% Anglo, and 2% other) revealed that the monolingual group at the Mid-Atlantic site was predominantly African American. Given that calcified achievement discrepancies have been documented not just between bilinguals and monolinguals, but also between students of color and Anglo students, this particular site difference, which added to the diversity and inclusiveness of the sample, was noteworthy and led to broader disaggregation of results (see Table 1).

Measures

Students were assessed in the fall (Time 1) and spring (Time 2) of the 2009–2010 academic year on English language measures. Bilingual students were assessed once in the winter (Time 1.5) of 2010 on Spanish language measures.

Table 1 N (percentage of sample) of students by site disaggregated by grade and ethnolinguistic status

Demographic indicators	Mid-Atlantic	Northeast	Total
Grade[a]			
2	77 (26.2)	29 (9.9)	106 (36.1)
3	65 (22.1)	29 (9.9)	94 (32.0)
4	63 (21.4)	31 (10.5)	94 (32.0)
Ethnolinguistic status[a]			
Anglo	11 (3.8)	36 (12.4)	47 (16.0)
Af. Am	95 (32.6)	6 (2.1)	101 (34.4)
Other	6 (2.1)	8 (2.7)	14 (4.8)
Bilingual ELL	55 (18.9)	10 (3.4)	65 (22.1)
Bilingual non-ELL	36 (12.4)	28 (9.6)	64 (21.8)
Grade × Ethnolinguistic status[b]			
Grade 2			
Anglo	6 (5.7)	13 (12.3)	19 (17.9)
Af. Am	38 (35.8)	3 (2.8)	41 (38.6)
Other	2 (1.9)	1 (1.0)	3 (2.9)
Bilingual ELL	24 (22.6)	4 (3.8)	28 (26.4)
Bilingual non-ELL	6 (5.7)	8 (7.5)	14 (13.2)
Grade 3			
Anglo	4 (4.3)	12 (12.8)	16 (17.1)
Af. Am	23 (21.7)	1 (1.1)	24 (25.5)
Other	3 (3.2)	4 (4.3)	7 (7.4)
Bilingual ELL	17 (18.1)	0 (0)	17 (18.1)
Bilingual non-ELL	18 (19.1)	12 (12.8)	30 (31.9)
Grade 4			
Anglo	1 (1.1)	11 (11.7)	12 (12.8)
Af. Am	34 (36.2)	2 (2.1)	36 (38.3)
Other	1 (1.1)	3 (3.2)	4 (4.3)
Bilingual ELL	14 (14.9)	6 (6.4)	20 (21.3)
Bilingual non-ELL	12 (12.8)	8 (8.5)	20 (21.3)

N = 291 for ethnolinguistic status as 3 students were missing relevant data

[a] Percentage derived from total sample

[b] Percentage derived from grade level sample

Word identification

The Woodcock-Muñoz Language Survey-Revised (WMLS-R; Woodcock et al., 2005) Letter-Word Identification subtest was used to capture this construct. On this measure, students were presented a list of real Spanish or English words ordered by increasing difficulty until 6 consecutive items were read incorrectly. Form A was administered to students. Raw scores were used for all analyses. The internal reliability of this subtest is .98 for 8-year-old children and .96 for 11-year-olds (Woodcock et al., 2005). Assessment was in English only. Raw scores were used in all analyses.

Breadth of vocabulary

The Woodcock-Muñoz Language Survey-Revised (WMLS-R; Woodcock et al., 2005) Picture Vocabulary subtest was used to capture breadth of expressive vocabulary. In this task, students were shown pictured items ordered by increasing difficulty and were asked to say aloud the names of each picture. Testing was discontinued after a student missed 6 consecutive items. Form A was administered to students. Raw scores were used for all analyses. The internal reliability for children 8 and 11 years old on the picture vocabulary test is .90 and .92 respectively (Woodcock et al., 2005). Assessments were conducted in English and Spanish. Raw scores were used in all analyses.

Morphology

The *Extract the Base* test (Anglin, 1993; August, Kenyon, Malabonga, Louguit, & Caglarcan, 2001; Carlisle, 1988) was individually administered to all students to evaluate awareness of derivational morphology. The test required a student to extract the base from a derived word (e.g., *farm* from *farmer*) when an examiner read aloud a target word (e.g., *farmer*) along with a contextual sentence (e.g., *My uncle works on a* ____). Students had worksheets showing the target words and sentences so they were able to follow the reading aloud of the prompt. Students then wrote the appropriate response in the blank area. Scores were calculated on a 0–2 coding scheme, where 0 indicated an incorrect response, 1 indicated a misspelling but phonologically plausible response (e.g., *proced* instead of *procede*), and 2 indicated a correctly spelled response. August et al. (2001) report Rasch-based reliability at .98. Assessments were conducted in English only. Raw scores were used in all analyses.

Semantics

The Clinical Evaluation of Language Fundamentals (CELF; Semel, Wiig, & Secord, 2003) Word Classes 2 subtest was used to measure awareness of semantic relations. In this task, students were read aloud four words, two of which were semantically related (e.g., *school*, *cake, street, teacher*). Students indicated which two words were semantically related. Testing was discontinued once a student missed 5 consecutive items. Test–retest reliability as indicated in the CELF manual ranges from .83 to .91 for children ages 7.0–9.11. The manual also reports internal consistency for these same ages as .73–.80. Assessments were conducted in English only. Raw scores were used in all analyses.

Syntax

The CELF Formulated Sentences subtest was used to measure this construct. On this task, students were shown a picture and given a target word that was to be used in a single sentence that described the picture. For example, a sample target word on this measure is *children*, and is accompanied by a picture of two children playing a video game. One common response to this prompt is the utterance, *The children are playing a video game*. Scores were calculated on a 0–2 coding scheme in which 0 indicated any of the following: Incomplete sentence; complete sentence with two or more deviations in syntax or semantics; complete sentence that is not meaningful; failure to use the stimulus word; failure to reference the stimulus picture. A response that received a 1 was a complete sentence that demonstrated correct structure and had only one or two deviations in syntax or semantics. Finally, a score of 2 represented a complete sentence that was semantically and syntactically correct and used a correct logical structure that was, meaningful, complete, and grammatical. Testing was discontinued after a student responded incorrectly (i.e., scored 0) on 5 consecutive items. Assessments were conducted in English and Spanish. Note that though the Formulated Sentences subtest was used primarily as a measure of syntax, it also captures semantics and morphology because it asks students to develop a morpho-synactically correct and meaningful sentence to get full credit for each particular item. Therefore, we used the Formulated Sentences subtest to indicate Spanish vocabulary depth. Test–retest reliability as reported in the CELF manual is .74–.79 for children ages 7.0–9.11 and internal consistency is .80–.82 for these same ages. Raw scores were used in all analyses.

Reading comprehension 1

The Woodcock-Muñoz Language Survey-Revised (WMLS-R; Woodcock et al., 2005) Passage Comprehension subtest was used to capture this construct. Form A was used in the fall and Form B was used in the spring. In this measure, students silently read cloze passages in order of increasing difficulty and produced an oral response to an unfinished sentence. The examiner then marked the response as correct or incorrect. The internal reliability of the passage comprehension assessment for children between 8.0 and 11.0 years old is .81–.91 (Woodcock et al., 2005). Because the WMLS Passage Comprehension subtest was used to anchor the comprehension factor score discussed below, W-scores were used for all analyses to provide for easier interpretation of the results. The choice of the W score for use in analysis will be discussed later.

Reading comprehension 2

The Gates–MacGinitie Reading Achievement Test (Forms S and T; MacGinitie, Maria, & Dreyer, 2002), was administered as a test of students' reading comprehension. This test has been used nationally and possesses strong psychometric properties. Test–retest reliability coefficients of the Gates MacGinitie are above .90 in fourth grade (MacGinitie et al., 2002). Level 3 was used for Grade 3, Level 4 for Grade 4, and Level 5 for Grade 5. Form S was administered in the fall, and Form T was administered in the spring. Raw scores were used for all analyses.

Reading comprehension 3

The Test of Sentence Reading Efficiency and Comprehension (TOSREC; Wagner, Torgeson, Rashotte, & Pearson, 2010) was administered as the third comprehension indicator. In this assessment student have 2 min in which they read a series of short, single sentence passages that they must rate as true or false (e.g., *A doorbell is used to sleep at night*). The authors report excellent alternate-form reliability ($r = .93$) and strong concurrent criterion-related validity ($r = .87$–$.89$). Raw scores were used for all analyses.

Analyses

We employed structural equation modeling (SEM) to answer the proposed research questions. SEM is an analytic method for modeling linear relations among latent variables and defines two regression models that play distinctive roles in formulating various substantive problems: (1) a measurement model that defines the relation of the latent variables (factors) to their observed variable indicators, and (2) the structural model, which defines the effects of latent and/or observed predictor variables on the latent outcome variable. For the purposes of this study, we modeled change in latent reading comprehension using a latent difference score (LDS; McArdle, 2001) model with time-invariant covariates. Figure 1 displays the prototypical latent difference score model used in subsequent analyses.

Shown in squares in Fig. 1 are the same three observed variables indicators of reading comprehension gathered at two occasions. The circles in the figure represent the latent factors. The first-order factors, RC-1 and RC-2, represent reading comprehension at Time 1 and Time 2, while the second-order factors, I and S, represent the intercept (at Time 1) and Time 1–Time 2 change factors, respectively. Other parameters in the model are the variable intercepts (τ), the structure coefficients (λ) that relate each RC factor to its measured variable indicators, and the means of the growth factors (μ_α and μ_β). Straight arrows represent direct effects while curved two-headed arrows represent variances and covariances. In a LDS model, structure coefficients, measurement error/uniqueness, and variable intercepts can be estimated as well as the fitted variance and covariance associated with the two time points. Germane to this study, the primary goal of our analysis is to estimate regression coefficients (γ) measuring the direct effects of corresponding covariates on the change in latent reading comprehension over time. Before change over time can be assessed, however, the stability or invariance of measurement properties of latent reading comprehension must be determined. This is an important consideration with any time-dependent latent variable model because an implicit assumption is that any growth in latent reading comprehension reflects true change in the underlying phenomena and is not due to change that may occur in the measurement model.

A general strategy used in testing measurement invariance of the construct over time is to fit a confirmatory factor analysis model to data at each time and evaluate whether the form of the observed-latent variable relation is the same (i.e., a *measurement model*). For the LDS model proposed here, a continuum of invariance

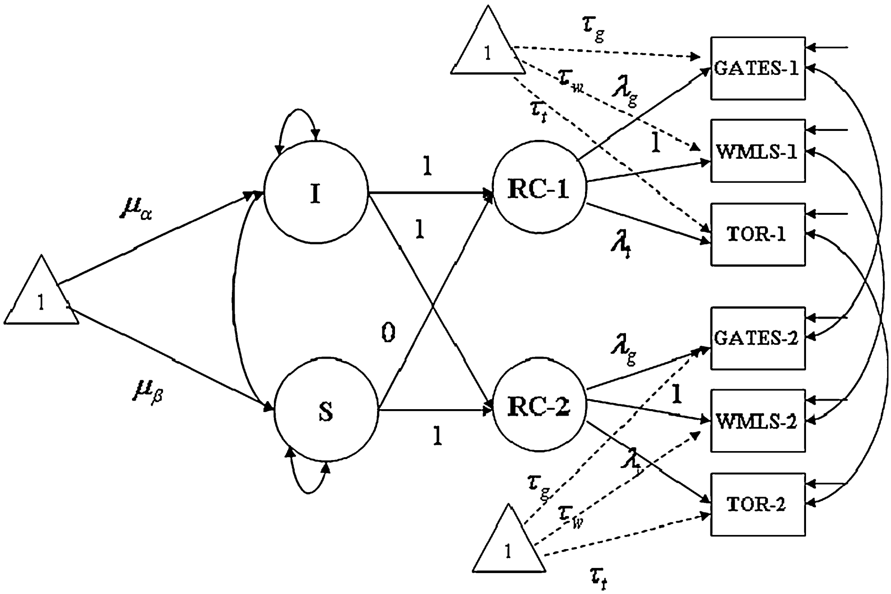

Fig. 1 Proposed Latent Difference Score model for assessing reading comprehension intercept (*I*) and change (*S*) over time using within- and between-group predictors

stringency was established by applying equivalence constraints to the structure coefficients, intercepts, and error variances of the measured variables.

In establishing the measurement model, we began with an unconstrained model and sought to establish whether the measurement model: (a) showed a consistent factor structure for comprehension across Times 1 and 2; (b) showed a consistent intercept across time points; and (c) showed consistent variation in comprehension across Times 1 and 2. These competing models were tested against one another to determine whether the characteristics of (a), (b), or (c) above most appropriately characterized the measurement model. See Fig. 2 for the proposed measurement model.

Once an appropriate measurement model was established, we next tested a structural model that analyzed the effects of grade (2, 3, 4) and language status (monolingual, bilingual ELL, bilingual non-ELL) on initial status of and change in reading comprehension (Research Question 1). We followed this step by creating a baseline reading model (i.e., decoding and vocabulary breadth) and then tested whether the three vocabulary depth variables (morphology, semantics, syntax) made unique contributions to intercept and Time 1–Time 2 change in English reading comprehension (Research Question 2). Finally, we disaggregated the bilingual (non-ELL and ELL) students and added Spanish vocabulary breadth and depth variables to assess whether Spanish proficiency made a unique contribution to intercept and change in English reading comprehension once English predictors were controlled (Research Question 3). See Fig. 3 for the proposed structural model.

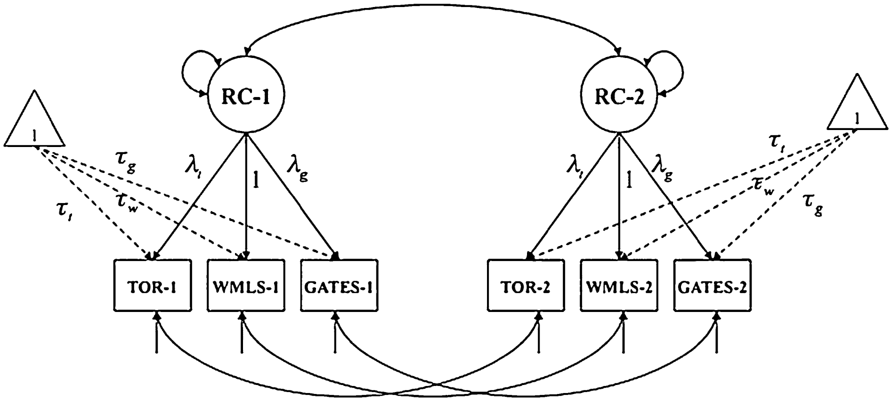

Fig. 2 Measurement model for Time 1 and Time 2 reading comprehension. *GATES* Gates Mac-Ginitie, *WMLS* Woodcock-Muñoz Language Survey, *TOR* Test of Sentence Reading Efficiency and Comprehension

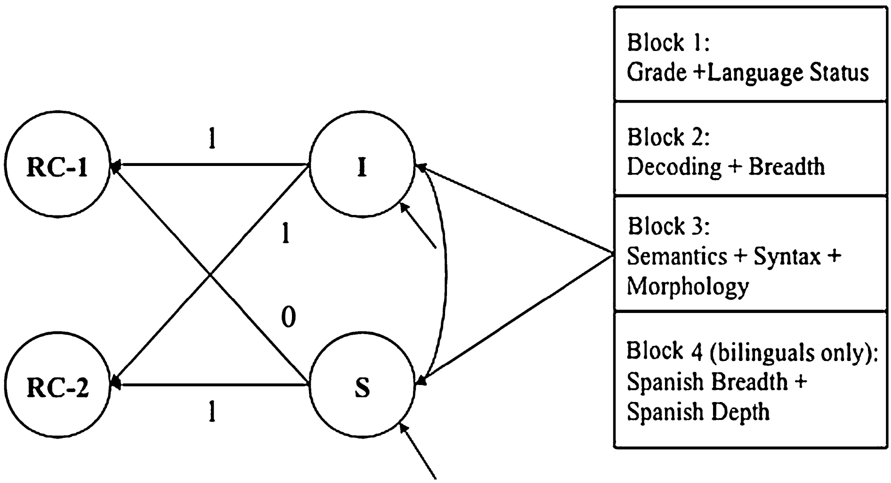

Fig. 3 Structural model displaying effects of depth predictors and Spanish proficiency controlling for: (1) Grade and language status; (2) Decoding and breadth of vocabulary. *RC* Reading Comprehension, *I* Intercept, *S* Change. Numbers 1 and 0 represent fixed factor loadings for linear growth: 1 for the intercept and 0–1 for the unit of change between Time 1–Time 2

We used raw scores for all reading comprehension and predictor measures except for the Woodcock-Muñoz Passage Comprehension subtest, for which we used W scores. The decision to equate factor scores to a W-distribution allowed for an interpretable growth metric for the major outcome of the study. The W score is a transformation of the Rasch scale with equal intervals. The W scale is centered on a value of 500 for the average performance of 10 year olds, and the distance between two points on the scale has the same interpretation regardless of age or grade level. Considering the range of ages and grade levels in this sample, using the W scale for the Passage Comprehension subtest to anchor our reading comprehension construct allowed for easier interpretation of the effect of independent variables on our latent dependent variable.

Results

Descriptive statistics

Table 2 displays the descriptive statistics for the sample, with data disaggregated by language status. Table 3 displays the results of one-way ANOVA testing on all outcome and predictor measures. Omnibus F-tests revealed significant differences by group, save for performance on the TOSREC comprehension measure. Tukey's pairwise procedure was used posteriori to test for group mean differences. In every instance but one, the monolingual and bilingual non-ELL groups were not significantly different from one another, but both were significantly different from the bilingual ELL group. Only on the measure of vocabulary breadth were the monolinguals significantly different from the bilingual non-ELL group, who were also significantly different from the bilingual ELL group (all *t*s > 1.96, all *p*s $< .05$). Among the bilingual groups, performance was comparable on the Spanish vocabulary depth measure, but there was a significant difference between the ELL and non-ELL groups on the measure of Spanish vocabulary breadth. However, it should be noted that performance overall on the Spanish vocabulary breadth measure was quite low for the bilingual non-ELLs and bilingual ELLs (11th and 14th percentiles, respectively).

Given the documented achievement gap between racial/ethnic groups, we also analyzed monolingual performance by the Anglo-American, African-American, and other ethnicity categories. Table 4 presents these descriptive data, and Table 5 presents the results of one-way ANOVA testing with Tukey's pairwise procedure also being used posteriori to test for group mean differences.

There were sharp discrepancies by race/ethnicity. In every instance but one, pairwise contrasts indicated that Anglo students were significantly different from their African American counterparts, while the Other group was not significantly different from either group. On the Gates MacGinitie at Time 1 the Anglo and Other groups were comparable with each other and both were significantly different from the African American group (all *t*s > 1.96, all *p*s $< .05$). Given the differences by language and race/ethnicity, we created an ethnolinguistic variable to use in analyses that disaggregated the monolingual group by race/ethnicity and the bilingual group by ELL/non-ELL. Thus, the ethnolinguistic variable included 5 categories: Anglo monolingual, African-American monolingual, Other monolingual, Latino/a bilingual non-ELL, and Latino/a bilingual ELL.

Table 6 displays the correlations between the measured variables, showing relatively strong and significant associations between most of the English variables. Associations between the TOSREC and other English variables were the least strongly associated. The two Spanish predictors were strongly associated with one another ($r = .645$, $p < .001$), and interestingly the Spanish vocabulary depth

Table 2 Reading variables for total sample and language status sub-groupings

English reading[a]	Total sample (n = 294)		Monolingual (n = 162)		Bilingual/non-ELL (n = 64)		Bilingual/ELL (n = 65)	
	Raw score mean (SD)	Mean percentile (SD)	Raw score mean (SD)	Mean percentile (SD)	Raw score mean (SD)	Mean percentile (SD)	Raw score mean (SD)	Mean percentile (SD)
GATES T1	23.5 (9.14)	38.09 (25.39)	25.03 (9.80)	42.72 (27.33)	24.64 (8.12)	41.34 (21.93)	18.76 (6.53)	23.92 (17.33)
GATES T2	26.01 (8.98)	35.83 (24.86)	26.85 (9.32)	38.56 (26.17)	27.38 (9.30)	40.40 (25.16)	22.58 (6.80)	24.49 (16.85)
WMLS-PC T1	481.14 (20.52)	43.75(25.42)	485.66 (16.65)	50.67 (25.10)	482.27 (16.18)	39.86 (23.96)	467.87 (17.44)	29.23 (20.59)
WMLS-PC T2	486.45 (15.36)	42.37 (22.25)	488.86 (15.80)	47.36 (24.24)	488.26 (13.22)	40.89 (18.34)	478.37 (13.65)	31.02 (15.40)
TOSREC T1	21.20 (8.36)	n/a	21.55 (8.28)	n/a	21.10 (7.90)	n/a	20.18 (8.98)	n/a
TOSREC T2	25.20 (8.60)	n/a	25.52 (8.40)	n/a	24.81 (8.80)	n/a	24.75 (9.0)	n/a
English predictors[b]								
Decoding	43.77 (10.26)	59.60 (29.02)	45.58 (9.62)	65.66 (27.30)	42.21 (9.35)	60.78 (28.82)	37.48 (10.48)	42.22 (27.13)
Breadth	30.91 (4.70)	52.75 (27.40)	32.89 (3.73)	64.89 (22.93)	30.50 (3.45)	48.03 (22.82)	25.97 (4.47)	24.72 (20.16)
Syntax	34.68 (9.89)	n/a	37.55 (9.22)	n/a	34.68 (8.89)	n/a	27.17 (8.63)	n/a
Semantics	7.33 (3.03)	36.18 (25.31)	8.02 (3.18)	42.01 (27.29)	7.50 (2.55)	34.48 (20.97)	5.35 (2.16)	20.28 (16.65)
Morphology	31.37 (1.43)	n/a	33.18 (10.84)	n/a	32.73 (10.55)	n/a	25.52 (11.90)	n/a
Spanish predictors[c]								
Breadth	n/a	n/a	n/a	n/a	19.98 (8.41)	10.51 (17.55)	24.47 (6.07)[d]	13.93 (15.56)
Syntax	n/a	n/a	n/a	n/a	20.26 (9.98)	n/a	21.51 (8.89)	n/a

GATES Gates-MacGinitie, *WMLS-PC* Woodcock-Muñoz Language Survey-Passage Comprehension, *TOSREC* Test of Sentence Reading Efficiency and Comprehension. Three students missing ethnolinguistic status data. WMLS-PC data are presented as W-scores as per the described analytic approach

[a] Measures administered at both Times 1 and 2

[b] Measures administered at Time 1 only

[c] Measures administered at Time 1.5 only

Table 3 Summary of One-Way ANOVA results for language status

	F	*df*	*p*
English reading			
GATES T1	12.52	2,285	.000
GATES T2	6.12	2,275	.003
WMLS-PC T1	18.91	2,286	.000
WMLS-PC T2	11.24	2,271	.000
TOSREC T1	.734	2,291	.481
TOSREC T2	.266	2,290	.766
English predictors			
Decoding	16.19	2,286	.000
Breadth	72.18	2,285	.000
Syntax	30.13	2,285	.000
Semantics	19.79	2,287	.000
Morphology	11.81	2,290	.000
Spanish predictors			
Breadth	7.6	1,119	.001
Syntax	.496	1,113	.483

See means and standard deviations in Table 2

variable showed some significant cross-linguistic associations with English decoding, syntax, and morphology as well as with Woodcock Passage Comprehension at Times 1 and 2. These correlations were in the mild-to-moderate range, from .267 to .306.

Research Question 1: what is the nature of beginning levels of reading comprehension and its change over one academic year? What are the effects of grade and ethnolinguistic status?

The measurement model

Figure 2 displays the tested measurement model in which we combined the 3 reading comprehension indicators onto single latent comprehension factors at Times 1 and 2. We used multiple indicators of model fit in assessing comparisons across competing models. The likelihood ratio test statistically compares the fit of the more restricted model with those of a comparable model without constraints (Brown 2006). Aside from the likelihood ratio test, other goodness-of-fit statistics have been recommended to test longitudinal measurement invariance (see e.g., Byrne & Stewart, 2006). These fit statistics include the comparative fit index (CFI, Bentler, 1990), root mean square error of approximation (RMSEA, Steiger, 1990), and standardized root mean square residual (SRMR, Steiger, 1990). CFI ranges from 0 to 1 with a larger value indicating better model fit. Acceptable model fit is indicated by a CFI value of 0.90 or greater (Hu & Bentler, 1999). RMSEA is related to residuals in the model. RMSEA values range from 0 to 1 with a smaller RMSEA value indicating better model fit. Acceptable model fit is indicated by an RMSEA

Table 4 Reading variables for monolingual students disaggregated by ethnicity (n = 162)

	Anglo (n = 47)		African American (n = 101)		Other (n = 14)	
	Score	Percentile	Score	Percentile	Score	Percentile
English reading[a]						
GATES Time 1	29.70 (9.16)	56.52 (26.83)	22.76 (8.98)	36.01 (24.33)	24.50 (12.15)	41.36 (32.89)
GATES Time 2	31.15 (7.61)	51.11 (23.60)	24.44 (8.93)	31.18 (23.66)	27.36 (12.05)	43.09 (34.16)
WMLS-PC Time 1	492.38 (14.53)	63.21 (22.12)	481.39 (16.39)	44.66 (24.34)	492.00 (16.61)	54.86 (25.28)
WMLS-PC Time 2	497.80 (14.99)	62.15 (23.33)	484.32 (14.80)	41.20 (22.44)	487.73 (12.33)	38.09 (17.69)
TOSREC Time 1	24.74 (8.70)	n/a	20.47 (7.65)	n/a	21.00 (8.73)	n/a
TOSREC Time 2	29.09 (8.08)	n/a	23.73 (7.65)	n/a	25.62 (10.64)	n/a
English predictors[b]						
Decoding	49.28 (7.78)	76.74 (21.86)	43.17 (9.72)	60.31 (28.12)	48.93 (9.49)	65.79 (28.65)
Breadth	35.11 (3.76)	75.87 (24.37)	31.85 (3.37)	59.79 (20.47)	32.64 (3.08)	64.11 (24.05)
Syntax	42.79 (8.01)	n/a	34.90 (8.91)	n/a	39.14 (6.92)	n/a
Semantics	9.78 (2.66)	55.45 (25.47)	7.14 (3.08)	35.96 (26.46)	7.93 (2.43)	41.00 (25.51)
Morphology	37.63 (9.14)	n/a	30.71 (11.20)	n/a	34.93 (8.13)	n/a

GATES Gates-MacGinitie, *WMLS-PC* Woodcock-Muñoz Language Survey-Passage Comprehension, *TOSREC* Test of Sentence Reading Efficiency and Comprehension

[a] Measures administered at both Times 1 and 2

[b] Measures administered at Time 1 only

[c] Measures administered at Time 1.5 only

value of 0.08 or less (Hu & Bentler, 1999). We used multiple sources of model fit to assess the level of measurement invariance to which our data conformed.

Table 7 displays the fitting process for the measurement model. Model 1 shows the results for the unconstrained model while Model 2 constrained the factor loadings for the indicator variable to remain the same at each measurement occasion. We used a threshold of $p < .05$ as a significance cutoff value. Comparing Models 1 and 2, a likelihood ratio test resulted in a non-significant result: ($\chi^2(2) = 4.8$, $p = .067$) indicating that the latent variable has the same impact on the indicators over time. For Model 3, in addition to retaining the equality constraints on the factor loadings, the intercepts for each indicator were constrained to be equal across time to test whether shifts in the mean of an indicator reflected substantive changes in the latent construct or simply changes in the intercept of the indicator. While the likelihood ratio test resulted in a model that was statistically

Table 5 Summary of One-Way ANOVA results for ethnicity

	F	*df*	*p*
English reading			
GATES T1	8.61	2,156	.000
GATES T2	9.01	2,150	.000
WMLS-PC T1	9.56	2,160	.000
WMLS-PC T2	12.96	2,149	.000
TOSREC T1	4.59	2,161	.012
TOSREC T2	7.01	2,159	.001
English predictors			
Decoding	8.18	2,160	.000
Breadth	14.0	2,159	.000
Syntax	14.05	2,161	.000
Semantics	13.16	2,160	.000
Morphology	7.21	2,160	.000

See means and standard deviations in Table 4

significant ($\chi^2(3) = 7.8$, $p = 0.05$), this was a borderline result and, combined with the other reported fit indices for Model 3 in Table 3 (i.e., RMSEA = 0.082, CFI = 0.948, SRSR = 0.082) the intercepts were deemed to be invariant over time. In Model 4, however, setting the variance equal over time produced a clearly significant likelihood ratio test and much poorer fit. It has been suggested (e.g., Sayer & Cumsille, 2001) that factorial invariance where error variances are constrained to be equal is unlikely to hold due to the fact that heterogeneous variance across time is often observed when modeling change and because the random error component of uniqueness may vary in magnitude with the level of the manifest variable. Model 3 was thus used to estimate intercept and change in latent reading comprehension.

Table 8 describes: (a) the model's intercept and Time 1 to Time 2 change; (b) amount of variance in both intercept and change; and (c) the covariance between intercept and change. Results suggest that the intercept and change score were significantly different from 0 such that, on average, students began with a W-score of 480.66 and gained an average of 5.91 W-units between fall and spring. Using W-score conversions, this equated to beginning the fall at a 3.1 grade level and gaining .5 of a grade level over the course of the year. Given the grade distribution of the sample (i.e., grades 2, 3, and 4) the grade level attainment was reasonable. While the change estimate was significant, it indicated just a half-year's growth (i.e., 4.5 months) in approximately 7-months' time elapsed between Time 1 and Time 2. However, this analysis also revealed a lack of significant variation in Time 1–Time 2 change (though variation was significant for the intercept). Finally, the covariance between intercept and change was negative and rather high (−15.07) yet non-significant ($r = -0.56$, $p = .13$), indicating that those students who showed lower Time 1 reading performance were more likely to show greater change at Time 2.

Table 6 Correlations among the observed variables in English and Spanish

	1	2	3	4	5	6	7	8	9	10	11	12
English measures												
1. GATES T1	–											
2. GATES T2	.704***	–										
3. WMLS-PC T1	.461***	.449***	–									
4. WMLS-PC T2	.537***	.524***	.678***	–								
5. TOSREC T1	.249***	.110	.182**	.145*	–							
6. TOSREC T2	.279***	.282***	.321***	.264***	.463***	–						
7. Decoding	.594***	.543***	.757***	.704***	.212***	.274***	–					
8. Breadth	.523***	.43***	.63***	.589***	.198**	.200**	.634***	–				
9. Syntax	.572***	.516***	.644***	.624***	.157**	.208***	.678***	.720***	–			
10. Semantics	.502***	.425***	.554***	.553***	.269***	.325***	.578***	.625***	.634***	–		
11. Morphology	.518***	.43***	.633***	.618***	.178**	.208***	.785***	.619***	.642***	.574***	–	
Spanish measures												
12. Breadth	−.020	−.071	.063	.142	.15	.069	.114	−.071	−.046	.049	.166	–
13. Depth	.038	.077	.306**	.301***	.013	.070	.267**	.058	.271**	.146	.280**	.645***

* $p < .05$; ** $p < .01$; *** $p < .001$

Table 7 Results from fitting the measurement model for time invariance

Model description	−2 ln$L$	# of parameters	Model comparison	D [($u - v$)df]	p	RMSEA	CFI	SRMR
1: Unconstrained	12440.6	22				.076	.985	0.023
2: $\lambda_{t1} = \lambda_{t2}$	12445.2	20	2 versus 1	5.4 (2)	.067	.080	.977	0.073
3: $\lambda_{t1} = \lambda_{t2}$ $\mu_{t1} = \mu_{t2}$	12453.1	17	3 versus 2	7.9 (3)	.050	.082	.948	0.082
4: $\lambda_{t1} = \lambda_{t2}$ $\mu_{t1} = \mu_{t2}$ $\sigma^2_{t1} = \sigma^2_{t2}$	12579.7	14	4 versus 3	126.6 (3)	<.001	.121	.791	0.121

Table 8 Assessing means, variance, and covariance among intercept and change for the latent reading construct

	Estimate	SE	t	p
Means				
Intercept	480.675	1.61	299.02	.000
Change	5.912	0.81	7.340	.000
Variance				
Intercept	148.45	24.92	5.96	.000
Change	4.79	8.03	.60	.55
Covariance				
Intercept with change	−15.07	9.89	−1.52	.13

r of intercept and change is −.56

Table 9 Establishing effects of grade on intercept and change of reading comprehension

Parameter	Est	SE	Est./SE	p
Sample intercept	481.09	1.8	261.1	<.001
Grade 2 versus 4	−5.80	3.1	−1.9	.057
Grade 3 versus 4	3.66	2.4	1.5	.134
Sample change	6.34	1.2	5.2	<.001
Grade 2 versus 4	0.88	1.6	0.6	.581
Grade 3 versus 4	−0.25	1.6	−0.2	.874

Referent group is Grade 4 students

Time 1 to Time 2 change for the entire sample was characterized by the slight diminishing of variation around the mean from Time 1 to Time 2. However, the likelihood ratio test of a model where the variances were freely estimated versus a model in which the variances were constrained to be equal was not statistically significant ($\chi^2(1) = 3.46$, $p = 0.07$), thus the variances of reading comprehension between Times 1 and 2 should be considered the same.

The role of grade and ethnolinguistic status

Tables 9 and 10 show the results of the analysis of intercept and change in reading comprehension as a function of grade level (Table 9) and ethnolinguistic status

Table 10 Establishing effects of language status on intercept and change of reading comprehension

Parameter	Est	SE	Est./SE	*p*
Sample intercept	492.77	2.2	225.7	<.001
Af.Am. versus Anglo	−13.30	2.6	−5.1	<.001
Bil ELL versus Anglo	−22.55	3.1	−7.2	<.001
Bil NonELL versus Anglo	−11.01	2.8	−3.9	<.001
Other versus Anglo	−7.44	4.3	−1.7	.087
Sample change	5.37	1.4	3.7	<.001
Af.Am. versus Anglo	−0.82	2.6	−0.5	.633
Bil ELL versus Anglo	3.86	1.9	2.0	.047
Bil NonELL versus Anglo	1.33	1.9	0.7	.477
Other versus Anglo	−1.99	3.1	−0.6	.523

Referent group is Anglo monolinguals

Af.Am African American, *Bil ELL* Bilingual ELL, *Bil NonELL* Bilingual Non-ELL

(Table 10). There were no significant differences between the groups for intercept or change when comparing grade level. Only the differences between the intercepts of second and fourth grades were approaching significance (est. = −5.80, $p = .057$). Table 2 showed that the Grade 4 cohort had a much higher percentage of African American students than the Grade 3 cohort (38.3% vs. 25.5%). Indeed, the African American students in the sample showed reading profiles more akin to those of the bilinguals than the Anglo monolinguals, which may be driving some of these counter-intuitive results.

Table 10 displays the results for ethnolinguistic status effects on initial status and slope, with Anglo students as the referent group and compared with their African-American, bilingual ELL, and bilingual non-ELL counterparts. Anglos significantly outperformed all other groups on initial status, with African-Americans and non-ELL bilinguals performing comparably. The ELL children were performing substantially below all their non-ELL monolingual and bilingual peers. In contrast, change estimates were non-significant for all groups except the ELL students, whose average Time 1–Time 2 growth was 9.23 W-units, representing approximately 80% of a year's growth in approximately 7 months' time (or 78% of an academic year). This effect is illustrated in Fig. 4. In subsequent analyses, we controlled for these between-subjects differences.

Research Question 2: do indicators of vocabulary depth (morphology, semantics, and syntax) predict initial status and change in reading comprehension beyond decoding and breadth of vocabulary?

Table 11 shows the effect of the baseline model of word reading and vocabulary breadth along with the three vocabulary depth predictors on initial status and slope of reading comprehension, controlling the between-subjects effects of grade and

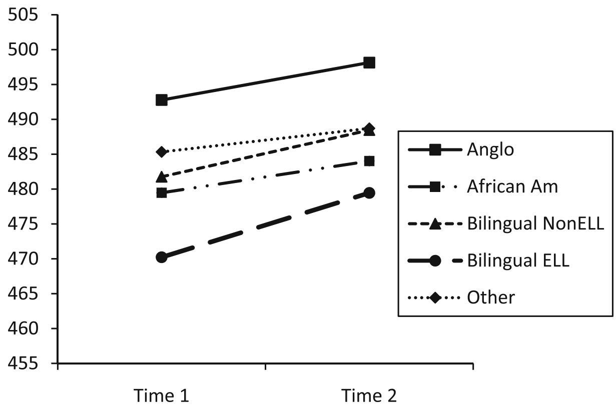

Fig. 4 Language status effects on reading comprehension growth from Time 1 to Time 2, as reported in Table 9

Table 11 Within-subjects depth model of reading comprehension controlling for grade and language status

Parameter	Intercept		T1–T2 Change	
	Est. (SE)	Standardized Est.	Est. (SE)	Standardized Est.
Conditional mean	476.56 (2.03)*		8.85 (2.33)*	
Baseline reading model				
Decoding	0.27 (0.03)*	0.58	−0.06 (0.04)	–
Breadth	0.17 (0.05)*	0.19	−0.10 (0.06)	–
Depth model predictors				
Semantics	0.89 (0.27)*	0.18	−0.27 (0.32)	–
Syntax	0.21 (0.09)*	0.13	0.04 (0.11)	–
Morphology	0.13 (0.09)	–	−0.05 (0.10)	–

Results take into account the between-subjects factors of grade level, language status, and ethnicity. Standardized estimates are reported for significant effects only as an indicator of effect size

* $p < .05$

ethnolinguistic status. Results revealed that semantics and syntax variables predicted initial status of reading comprehension, while morphological awareness did not. No vocabulary depth measure predicted change.

The strongest standardized effects on initial status were for word reading (.58) followed by vocabulary breadth (.19). Semantic awareness made an independent contribution at .18 while standardized estimate for syntactic awareness was .13. Cross-product variables that explored the interactions among the between- and within-subjects factors were non-significant, however, effects may have not have

Table 12 Within-subjects depth model of reading comprehension controlling for grade and language status for Spanish–English bilinguals only, testing for effects of Spanish language proficiency on English reading comprehension intercept and Time 1–Time 2 change

Parameter	Intercept		T1–T2 change	
	Est. (SE)	Standardized Est.	Est. (SE)	Standardized Est.
Conditional mean	466.57 (4.3)*		15.25 (4.2)*	
Baseline reading model				
Decoding	0.21 (0.09)*	0.43	−0.04 (0.05)	–
Breadth	0.45 (0.16)*	0.38	−0.20 (0.13)	–
English depth predictors				
Semantics	2.07 (0.61)*	0.31	−1.72 (0.57)*	−0.60
Syntax	0.05 (0.20)	–	0.01 (0.18)	–
Morphology	0.02 (0.16)	–	0.00 (0.12)	–
Spanish predictors				
Breadth	−0.01 (0.08)	–	−0.14 (0.21)	–
Depth	0.3 (0.22)	–	0.03 (0.07)	–

Results take into account the between-subjects factors of grade level and language status (ELL vs. non-ELL). Standardized estimates are reported for significant effects only as an indicator of effect size

* $p < .05$

been detectable due to decreasing degrees of freedom and a resultant loss of statistical power.

Research Question 3: among Spanish–English bilinguals, does Spanish language proficiency predict initial status and/or change in English reading comprehension?

Table 12 displays the results when the model was replicated among the Spanish–English bilinguals (ELL and non-ELL) with the addition of Spanish vocabulary breadth and depth. Comparing the conditional means with those displayed in Table 11 for the entire sample, the bilingual students' intercept was 10 W-points lower than for the entire sample (466.57 vs. 476.56) while the change estimate was markedly higher (15.25 W-points vs. 8.85 W-points), suggesting that the bilingual students in general were reading less proficiently in English but gaining more strongly. Also distinct from the whole sample model, the effect of English semantic awareness was quite strong while the effects of English syntax were absent. Further, English semantic awareness had a significant, negative, effect on Time 1–Time 2 change in English reading comprehension. This finding suggests that bilingual students with less English proficiency showed more Time 1–Time 2 change in English Reading Comprehension than students with stronger English proficiency. Finally, neither Spanish vocabulary depth nor vocabulary breadth made a significant contribution to explaining initial status or change in English reading comprehension. Spanish depth showed a positive, though non-significant, association with initial status ($\beta = .30$, $p = .16$).

Discussion

The results of the current analyses showed important contributions of vocabulary depth to reading comprehension, controlling for word identification and vocabulary breadth, as well as between-subjects factors related to grade and language status. In this study, we defined vocabulary depth as incorporating the metalinguistic skills of morphological awareness, semantic awareness, and syntactic awareness. Together, these skills allow language users to apply word knowledge flexibly across a variety of contexts. Results showed that both semantic and syntactic awareness were predictive of initial status of reading comprehension above and beyond word identification and vocabulary breadth. However, neither predicted change in reading comprehension across an academic year and morphology was not related to initial status or change in reading comprehension. Controlling for the other variables in the model, there were no differences by language background or ethnic group (i.e., monolingual Anglo and African-American children and bilingual ELL and non-ELL children). Thus, achievement differences by language background and ethnic group were erased when controlling for the linguistic variables in the model. For bilingual students alone, semantic awareness was the single strongest predictor of initial status beyond word identification and vocabulary breadth while Spanish language measures did not add to the prediction of English reading comprehension for the bilingual children in the sample. This result suggests that English language outcomes are driven primarily by English language rather than native language proficiency, at least for this group of Spanish speakers whose Spanish language proficiency was relatively underdeveloped and who received no Spanish literacy instruction through the participating schools.

Depth of vocabulary and reading comprehension

Previous researchers have highlighted the importance of metalinguistic awareness, which we refer to here as depth of word knowledge, to children's reading and cognition (Kuo & Anderson, 2010; Bialystok, 2006; Hakuta, 1986; Hakuta & Diaz, 1985). However, there is limited research that investigates multiple forms metalinguistic awareness across children from different language backgrounds. This study adds to the literature by investigating multiple facets of depth of word knowledge within one study.

Findings from this study diverged from findings of other researchers on the role of morphological awareness in reading comprehension. For example, Deacon and Kirby (2004) found significant associations between morphology and reading comprehension in children from 2nd through 5th grade, controlling for other reading-related skills. However, the measure used in the Deacon and Kirby (2004) study was administered orally and required students to respond orally whereas the morphological awareness measure used in this study (i.e., Extract the Base) was administered orally but required students to respond in writing, which taps not only morphological awareness but also orthographic knowledge. Thus, it was not surprising that in the current study morphological awareness was very strongly correlated with our measure of decoding ($r = .785$, $p < .001$). The overlap of variance between decoding and morphology may have altered the relationship between morphology and reading. Future research should investigate multiple measures of morphology to explore whether relationships between morphology and reading comprehension are stable when measurement error is removed. Of course, morphology, semantics, and syntax are interrelated so having measures of each of these forms of metalinguistic awareness in one model could also affect results. In this study, the measure of morphological awareness used was correlated .574 with the measure of semantics and .642 with the measure of syntax (ps $< .001$). The discrepancy between findings in previous research and the present study for morphology clearly warrants further investigation.

Unlike morphology, semantics was predictive of comprehension. In fact, the standardized strength of the semantic contribution to comprehension (.18) was almost identical to that of vocabulary breadth (.19), and among the bilingual students this relationship was even stronger (.31), though still assumed a lesser standardized effect when compared with decoding (.43) and vocabulary breadth (.38). Previous research with younger learners found that semantic awareness did not add to reading comprehension after controlling for vocabulary breadth (Tannenbaum et al., 2006). The findings presented here, however, are in line with the work of Ouellette (2006) and Proctor et al. (2009) who reported that semantic depth played a significant role in the reading comprehension of fourth grade students even after controlling for decoding and vocabulary breadth. However, given the sample size, we were unable to test interactions between depth measures and grade or language background. Thus, future research should also investigate differences in the contribution of semantics to reading comprehension by these between-subjects factors. Overall, however, this study suggests that it is important to consider not only how many words children know but also how well children understand connections between and among words across contexts.

Finally, syntax was also predictive of comprehension. Its standardized effect (.13) was smaller than that of semantics, but the finding that syntax was relevant in predicting comprehension is notable. Some research has found a relationship between syntactic awareness and reading comprehension (Mokhtari & Thompson, 2006; Nation & Snowling, 2000), and this study adds to that literature. Syntactic awareness relates to combining words in context to create grammatically correct and semantically meaningful sentences. Therefore, syntactic knowledge may affect facility with connected text more than semantic and morphological awareness (Cain, 2007). Returning to the *table* example, good syntactic knowledge would be present when a student understood the use of the word in various contexts: a) as a simple noun at home; and b) as an academic verb in science class. Advanced morphosyntactic knowledge would be present when that same student could derive the term *tabulate* in the context of describing what she did in constructing the table for her science report. Awareness of syntax, therefore, is important for students to understand how words are connected as they navigate the connected text they read in school.

Reading comprehension

In addition to investigating multiple facets of vocabulary depth, this study adds to the literature by investigating these facets in relation to a reading comprehension factor that combines multiple measures of reading comprehension that assess the construct in different ways. Given the complexity of the construct of reading comprehension (Snow, 2002), it is not surprising that different measures of reading comprehension show different relationships to other variables depending on the format and content of the measure (Cutting & Scarborough, 2006). In this study, we used three measures that assessed comprehension in different ways: (a) a cloze task that required students to read a sentence or paragraph and fill in the blank, (b) a multiple choice task that asked students to read passages and answer questions about those passages, and (c) a timed sentence verification task in which children decided whether or not a sentence made sense. By using a factor score combining these measures into one latent construct, we were able to capitalize on different aspects of comprehension and, simultaneously, minimize measurement error inherent in using only one measure to assess a skill.

In this study, the effect of the language and literacy measures was found on initial status, but not on change in reading comprehension across the school year. It is most likely that our measures did not predict change because there was insufficient variability in reading comprehension change from T1–T2 (as shown in Table 8). It may also be that once the effect on initial status was taken into account, there was no longer any predictive effect of the language and literacy measures on reading comprehension. One final explanation is that time (or instruction) may have an equal effect on all students regardless of their language and literacy, thereby creating limited variability in change in reading comprehension to explain.

Bilingualism and reading comprehension

Spanish language proficiency was not associated with English reading comprehension as has been demonstrated in previous research (i.e., Nakamoto et al., 2008; Proctor et al., 2006), in line with Mancilla-Martinez and Lesaux's (2010) findings. What also links this finding with that of Mancilla-Martinez and Lesaux (2010) is that none of the bilingual students in either study received any form of Spanish language instruction. By contrast, the students in the Nakamoto et al. (2008) and Proctor et al. (2006) studies all received some degree of Spanish language and literacy instruction. Again, the role of instruction is present in this non-finding, but with bilingual learners the instructional question is more complex. Quality of instruction is imperative for all learners, but for bilingual learners, there is the added question of language of instruction. Might the provision of quality heritage language instruction for bilingual learners support depth of vocabulary learning? Without wading into the tedious debate around bilingual education, future research should consider this question, specifically whether Spanish instruction serves as a gatekeeper for accessing cross-linguistic associations between morphology, semantics, and syntax among upper elementary and middle school learners.

Limitations and future directions

This study relied on mostly published, standardized, norm-referenced measures, which are quite often biased against children of color (Ladson-Billings, 2006; MacSwan & Rolstad, 2006); Valdés & Figueroa, 1994). While we made efforts to find measures that included students of color in the norming sample, the findings are most certainly limited by the measures we chose. Further, the present study did not investigate socio-economic and socio-emotional factors, including immigration experiences, trauma, and long-term exposure to institutionalized discrimination (Geronimous, 2003), that affect the achievement of linguistically and culturally diverse children. Future research in literacy and development must explore how these sociological forces affect children's language and literacy development.

An important limitation to the current research is the many cells into which children were categorized (language status, ethnicity, grade) and the effect on our ability to detect differences in Time 1–Time 2 change. Given the small samples of each group, such analyses were not possible. Thus, future research would do well to examine the developmental trajectories of children within these categories when sample sizes are larger and measurement occasions more frequent.

Additionally, this study did not explore the impact of instruction on children's developmental trajectory. Future research should specifically focus on instructional variables given work by other researchers that suggests its impact on growth in reading. For example, Connor, Morrison, and Underwood (2007) used hierarchical linear modeling to address this question in first and second grade English language arts classrooms. In this work, Connor et al. (2007) observed and coded language arts instruction taking into account a wide variety of instructional practices, which were characterized as: a) teacher-managed or student-managed instruction (TM or SM); and b) code-focused or meaning-focused instruction (CF or MF). The authors found that TM-CF instruction predicted growth in word reading, which interacted with students' entering reading skills. Research that investigates these instructional effects among upper elementary school-age children is needed to determine how teaching and learning interact as meaning making takes precedence over code breaking.

Conclusion

Comprehension involves not just reading words and knowing what they mean but how they are connected in language to make meaning. The results presented here open some important windows on the comprehension process for bilingual and monolingual students and the developmental and instructional approaches educators can take to understand reading achievement among diverse populations. That the vocabulary depth measures predicted English reading performance suggests that these may be used as screening for children at risk for reading problems and that these skills may be candidates for instruction. What remains unclear is the age, grade, or linguistic proficiency levels at which vocabulary depth instruction might be most efficiently leveraged. These are robust candidates for future research, most particularly among culturally and linguistically heterogenous groups of learners across the developmental spectrum.

References

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of Society Research in Child Development, 58*, (10, Serial No. 238).

August, D., Kenyon, D., Malabonga, V., Louguit, M., & Caglarcan, S. (2001). *Extract the base*. Washington, DC: Center for Applied Linguistics.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Beringer, V., Abbot, R., Nagy, W., & Carlisle, J. (2009). Growth in phonological, orthographic, and morphological awareness in grades 1–6. *Journal of Psycholinguistic Research, 39*(2), 141–163.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60*, 68–79.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*, 498–520.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications Inc.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling, 13*, 287–321.

Cain, K. (2007). Syntactic awareness and reading ability: Is there evidence for a special relationship? *Applied Psycholinguistics, 28*, 679–694.

Carlisle, J. F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics, 9*(3), 247–266.

Carlisle, J. F. (2007). Fostering morphological processing, vocabulary development, and reading comprehension. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 78–103). New York: Guilford.

Carlisle, J. F., Beeman, M. M., Davis, L. H., & Spharim, G. (1999). Relationship of metalinguistic capabilities and reading achievement for children who are becoming bilingual. *Applied Psycholinguistics, 20*, 459–478.

Child Trends DataBank. (2005). Dropout rates. Washington, DC: Author. Retrieved March 18, 2009, from www.childtrendsdatabank.org/pdf/1_PDF.pdf.

Connor, C. M., Morrison, F. J., & Underwood, P. S. (2007). A second chance in second grade: The independent and cumulative impact of first- and second-grade reading instruction and students' letter-word reading skill growth. *Scientific Studies of Reading, 11*(3), 199–233.

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*(3), 277–299.

Deacon, S. H., & Kirby, J. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics, 25*, 223–238.

Geronimous, A. (2003). Damned if you do: Culture, identity, privilege, and teenage childbearing in the United States. *Social Science and Medicine, 57*(5), 881–893.

Geva, E. (2006). Second-language oral proficiency and second-language literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language Minority Children and Youth* (pp. 123–140). Mahwah, NJ: Laurence Erlbaum Associates.

Gough, P. B., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.

Graves, M. (2006). *The vocabulary book: Learning and instruction*. New York: Teachers College Press.

Hakuta, K. (1986). Degree of bilingualism and cognitive ability in mainland Puerto Rican children. *Child Development, 58*(5), 1372–1388.

Hakuta, K., & Díaz, R. (1985). The relationship between bilingualism and cognitive ability: A critical discussion and some new longitudinal data. In K. E. Nelson (Ed.), *Children's language* (Vol. 5, pp. 319–344). Hillsdale, NJ: Lawrence Erlbaum.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

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

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.

Kieffer, M. J., & Lesaux, N. K. (2007). Breaking down words to build meaning: Morphology, vocabulary, and reading comprehension in the urban classroom. *The Reading Teacher, 61*(2), 134–144.

Kieffer, M. J., & Lesaux, N. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing: An Interdisciplinary Journal, 21*(8), 783–804.

Kuo, L. J., & Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist, 41*(3), 161–180.

Kuo, L., & Andreson, R. C. (2010). Beyond cross-language transfer: Reconceptualizing the impact of early bilingualism on phonological awareness. *Scientific Studies of Reading, 14*(4), 365–385.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational Researcher, 35*(7), 3–12.

Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology, 99*(4), 821–834.

Low, P., & Siegel, L. S. (2005). A comparison of the cognitive processes underling reading comprehension in native English and ESL speakers. *Written Language and Literacy, 8*(2), 207–231.

MacGinitie, W., MacGinitie, R., Maria, K., & Dreyer, L. (2002). *Gates-MacGinitie. Reading tests* (4th ed.). Itasca, IL: Riverside.

MacSwan, J., & Rolstad, K. (2006). How language tests mislead us about children's abilities: Implications for special education placements. *Teachers College Record, 108*(11), 2304–2328.

Mancilla-Martinez, J., & Lesaux, N. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of Educational Psychology, 102*(3), 701–711.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. H. C. du Toit, & K. G. Joreskog (Eds.), *Structural equation modeling: Present and future* (pp. 341–380). Lincolnwood, IL: Scientific Software.

Mokhtari, K., & Thompson, H. B. (2006). How problems of reading fluency and comprehension are related to difficulties in syntactic awareness skills among fifth graders. *Reading Research and Instruction, 46*(1), 73–94.

Nagy, W. E., Berninger, V. W., & Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle school students. *Journal of Educational Psychology, 98*(1), 134–147.

Nagy, W. E., & Herman, P. A. (1985). Incidental vs. instructional approaches to increasing reading vocabulary. *Educational Perspectives, 23*, 16–21.

Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2008). A cross-linguistic investigation of English language learners' reading comprehension in English and Spanish. *Scientific Studies of Reading, 12*, 351–371.

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research, 47*, 199–211.

Nation, K., & Snowling, M. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342–356.

National Center for Education Statistics. (2007). *National assessment of educational progress*. Washington, DC: Institute on Education Statistics.

Ordoñez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology, 94*(4), 719–728.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*, 554–566.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282–296.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven (Ed.), *Precursors of functional literacy* (pp. 189–213). Philadelphia, PA: John Benjamins.

Proctor, C. P., August, D., Carlo, M. S., & Snow, C. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English reading comprehension. *Journal of Educational Psychology, 98*, 159–169.

Proctor, C. P., Carlo, M. S., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology, 97*(2), 246–256.

Proctor, C. P., Uccelli, P., Dalton, B., & Snow, C. (2009). Understanding depth of vocabulary online with bilingual and monolingual children. *Reading and Writing Quarterly, 25*(4), 311–333.

Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179–200). Washington, DC: American Psychological Association.

Semel, E. M., Wiig, E. H., & Secord, W. (2003). *Clinical evaluation of language fundamentals*. San Antonio, TX: The Psychological Corporation.

Snow, C. E. (2002). *Reading for understanding toward a R & D program in reading comprehension*. Santa Monica, CA: RAND.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading, 10*(4), 281–298.

Valdés, G., & Figueroa, R. A. (1994). *Bilingualism and testing: A special case of bias*. Norwood, NJ: Ablex.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*(2), 217–234.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. (2010). *Test of silent reading efficiency and comprehension*. Austin, TX: PRO-Ed.

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics, 27*(4), 741–747.

Woodcock, R., Munoz-Sandoval, A., Reuf, M., & Alvarado, C. (2005). *Woodcock-Muñoz language survey-revised*. Itasca, IL: Riverside Publishing Company.