Academic Vocabulary and Reading Comprehension: Exploring the Relationships Across Measures of Vocabulary Knowledge

Abstract

This study examined the relationship between knowledge of academic vocabulary and reading comprehension in data contributed by 5,855 middle school students. Each student completed an academic vocabulary assessment, a standardized reading comprehension test, and one of four types of novel vocabulary-depth measures. *Multiword expressions* examined students' abilities to complete formulaic phrases. *Topical associates* items required students to identify a target word that was topically related to three others. The *hypernyms task* required students to identify the superordinate for each target word. The *definitions task* asked students to choose the definition of the target word. We modeled the relationship between performance on the reading comprehension task and each of the four types of assessments using a residual factors approach (Bentler & Satorra, 2000) with latent variables. Even though each depth measure tested exactly the same sets of words, we found that these measures had a differential impact on reading comprehension, with the definitions task explaining the largest portion of variance in reading comprehension beyond overall academic vocabulary. The knowledge of multiword expressions and topical associates—but not of hypernyms—also explained unique variance in reading comprehension even when controlling for academic vocabulary knowledge.

Keywords: academic vocabulary, vocabulary depth, assessment, adolescent

Academic Vocabulary and Reading Comprehension: Exploring the Relationships Across Measures of Vocabulary Knowledge

Word knowledge plays a multifaceted role in reading comprehension (Perfetti & Stafura, 2014). There are phonological, orthographic, syntactic, and semantic ways of knowing a word (Perfetti, 2007; Perfetti & Hart, 2002; Perfetti, Wlotko, & Hart, 2005), which together constitute lexical quality, a central component of skilled reading. Despite the widespread understanding that the semantic component of word knowledge is itself extremely rich, relatively few studies have examined the relationship between different aspects of semantic word knowledge and reading comprehension. The current study examined four dimensions of vocabulary knowledge using novel multiple-choice assessments of vocabulary depth (Deane et al., 2014). We tested whether performance on these measures explained reading ability, even controlling for student knowledge of academic vocabulary, measured with a 50-item synonym task.

Vocabulary and Reading Comprehension

Vocabulary knowledge and reading ability correlate across childhood (McKeown, Beck, Omanson, & Perfetti, 1983; Quinn, Wagner, Petscher, & Lopez, 2015; Snow, Porche, Tabors, & Harris, 2007; Tannenbaum, Torgesen, & Wagner, 2006; Wagner et al., 1997) and especially in adolescence (Cromley & Azevedo, 2007). There are several explanations for these strong correlations (Anderson & Freebody, 1981; Quinn et al., 2015), none of which are mutually exclusive. The *instrumentalist hypothesis* suggests a causal relationship with a direct impact of vocabulary knowledge on comprehension; the more words a reader knows in a passage, the better comprehension the reader will have. This instrumentalist perspective informs research that estimates the percentage of words that readers need to know in order to comprehend a passage

(Schmitt, Jiang, & Grabe, 2011) and the use of word frequency and dispersion estimates to determine which words are of high utility for reading (Coxhead, 2000).

However, instructional intervention studies show limited support for the instrumentalist perspective. Studies that test the effect of vocabulary training on comprehension and recall tasks find an interaction between targeted word instruction and recall of stories that use target words (Beck, Perfetti, & McKeown, 1982; McKeown et al., 1983). However, attempts to improve reading comprehension by teaching larger sets of instructional target words selected with reference to instrumentalist utility (Lawrence, Francis, Pare-Blagoev & Snow, 2017; Lawrence, Crosson, Pare-Blagoev & Snow, 2015; Lesaux, Keiffer, Kelley & Harris, 2014; Lesaux, Kieffer, Faller, & Kelley, 2010; Pany, Jenkins, & Schreck, 1982) have found no main effect, or only marginal effects, on general reading comprehension. Meta-analyses show that when measured with researcher-developed instruments, the effect of direct vocabulary instruction on passage comprehension is only moderate (Elleman, Lindo, Morphy, & Compton, 2009; Stahl & Fairbanks, 1986) and that standardized reading measures show no such effects.

A second, nonexclusive explanation for the strong correlation between vocabulary and reading comprehension is that knowing a word is actually part of a person's knowledge about the world, and the more a reader knows about the world, the better they are able to understand what they read. This noncausal perspective has been referred to as the *knowledge hypothesis* (e.g., Anderson & Freebody, 1981; Stahl & Nagy, 2007) and implies that if knowledge is the lynchpin, then knowing multiple meanings of a single word might be just as important as knowing several words. Certain words may also have more utility because of their generality or specificity or because they name causal or intertextual relationships that are key to reading comprehension, even if the words are relatively infrequent. But there is scant quantitative research on this

perspective, possibly because it is hard to usefully estimate semantic dimensions conveyed by words, whereas it is relatively easy to report frequency estimates from large corpora (e.g., Kučera & Francis, 1967).

A third noncausal and nonexclusive perspective on the relationship between vocabulary and reading comprehension is the *aptitude hypothesis*. It is well understood that general factors may be responsible for high correlations in cognitive skills (Spearman, 1904; Tucker-Drob, 2009), and verbal ability has been identified as a measure related to general intelligence (Carroll, 1941; Sternberg & Powell, 1983; Thurstone, 1938). The aptitude hypothesis emphasizes that vocabulary knowledge and reading comprehension are supported by an underlying general verbal aptitude; this underlying aptitude explains the correlation between the two. In the reading research literature, metalinguistic awareness has been identified as the ability to "reflect on and manipulate the structural features of spoken language" (Tunmer & Herriman, 1984, p. 136). Nagy (2007) explicitly referenced this skill as an explanation for individual differences in vocabulary knowledge and the strong relationship between vocabulary knowledge and reading comprehension. The aptitude hypothesis is not mutually exclusive of the others, nor does it preclude the possibility of rich reciprocal relationships between vocabulary knowledge and reading comprehension (Stanovich, 1986; Verhoeven, van Leeuwe, & Vermeer, 2011).

The speed and accuracy with which we can retrieve the meaning of a word may also explain the correlation between vocabulary knowledge and reading performance (LaBerge & Samuels, 1974; Stanovich, 1986). Although it takes both fluent recognition/decoding and language comprehension to access semantic knowledge (Gough & Tunmer, 1986), word fluency has consistently been the best predictor of lexical access (despite limitations; see Kuperman & Van Dyke, 2013), followed by phonological and orthographic features (Yap, Tan, Pexman, &

Hargreaves, 2011). However, even with careful controls, there are differences in performance on speeded lexical decision tasks based on purely semantic features, such as how many meanings a word has; words with more related meanings tend to be accessed more quickly, but words with more distinct meanings tend to be accessed more slowly (see Eddington & Tokowicz, 2015 for a review). Some studies have manipulated which meanings are presented in reading contexts and have demonstrated that negotiating the meaning changes of polysemous target words across sentences results in slower reading speeds (Foraker & Murphy, 2012). In one of the few intervention studies to examine these issues, McKeown and colleagues found improvement both in how quickly children completed semantic decision tasks with target words and how well they read texts that included them (McKeown et al., 1983). Again, this is not to suggest that this explanation is independent of the others. Indeed, there is evidence that both accuracy of word knowledge and speed of retrieval make independent contributions to concurrent text comprehension by children (Oakhill, Cain, McCarthy, & Nightingale, 2012; Richter et al., 2013).

Dimensions of Vocabulary Knowledge

The current study focuses on four understudied dimensions of vocabulary knowledge, as described below.

Multiword expressions. Many word bundles occur in texts more frequently than would be expected by chance (Biber, Conrad, & Cortes, 2004; Hyland, 2012). For instance, *aptitude for music* and *paralyzed with fear* are phrases that occur (relatively) frequently. Mutual information (MI) is a statistical measure used to determine which of the words in a phrase occur together frequently (Ellis, Simpson-Vlach, & Maynard, 2008; Manning & Schuetze, 1999). Not surprisingly, these formulaic sequences are processed efficiently by both native and nonnative

speakers (Conklin & Schmitt, 2008). Interestingly, for native speakers, a phrase's processability is best predicted by MI, whereas for nonnative learners, it is best predicted by the frequency of the formula (Ellis et al., 2008). Reading research in which text is manipulated to include or exclude multiword expressions shows that the occurrence of these expressions impacts comprehension, even controlling for the frequency of words used in the passages (Martinez & Murphy, 2011). These findings suggest that knowing which words are co-located with a target vocabulary item may be an important dimension of word knowledge, one that relates to efficient text comprehension. Surprisingly, there are few studies assessing this potentially important dimension of word knowledge.

Topical associates. Understanding how we parse similarities across words to establish categories is a foundational topic in the philosophy of language (Wittgenstein, 1958), possibly because this task implicates vocabulary knowledge, comprehension, and memory (Anderson et al., 1976; Anderson & Ortony, 1975; Halff, Ortony, & Anderson, 1976). Indeed, some researchers have understood vocabulary knowledge primarily as network building (Haastrup & Henriksen, 2000) and have even suggested that there is no difference between knowing a word well and having a rich lexical network related to that word (i.e., there is no distinction between vocabulary depth and breadth; Vermeer, 2001). Word associations themselves can be parsed in many ways, however. Jenkins (1970) used data from free word associations (with one stimulus and one response) to show that subjects use four common strategies to produce related words. The most common strategy is to look for words with the same level of detail, such as *salt* and *pepper.* Subjects also use collocation (*salt water*), superordination (an example of which would be the response of *dog* to the stimulus *poodle*), and synonymy to free associate word relationships.

The use of word-list recall tasks has greatly enhanced our understanding of lexical processing and word associations. Researchers have successfully induced false memories in subjects by providing them with lists of related words that lack a salient target (Roediger & McDermott, 1995). One explanation for the false memory is that while the lists are being presented, our brains trigger presented and nonpresented words, but we fail to monitor the source of the activation and so do not inhibit the false memory (Roediger & McDermott, 1999). An alternative view, *fuzzy trace theory*, suggests that the world knowledge of subjects influences their interpretation. In this view, subjects store phonological and semantic representations in parallel and form *gist traces* of semantically related words during storage (Brainerd & Reyna, 2005; Reyna & Kiernan, 1994). Significantly, it has been shown that poor comprehenders (ages 9–11) are less likely than strong comprehenders to experience false memories when confronted with semantically related words (Weekes, Hamilton, Oakhill, & Holliday, 2008). These studies demonstrate that the ability to deduce the topic of related words is complex and may be related to reading comprehension, but this hypothesis has not been tested with vocabulary assessments.

Hypernyms. A hypernym is a superordinate general term that subsumes a set of specific hyponyms. For instance, *dog* is a hypernym to *poodle*, *terrier*, and *mutt*. Collins and Quillian (1969) argued that our mental lexicon is stored in hypernym chains (animal > dog > poodle). They asked subjects to verify sentences like *a poodle is a kind of dog* and *a poodle is a kind of animal*. They found that subjects responded faster when the tested hypernym chains were shorter. These effects held even when associative frequency was held constant (Conrad, 1972). Interestingly, Shaeffer and Wallace (1970) and Wilkins (1971) found that similarity increases the difficulty of making negative judgements. Thus, *a canary is an ostrich* is judged more **slowly** than *a canary is a fish*. In a follow-up study, Johnson-Laird (1983) hypothesized that subjects

respond faster to adjacent higher-order hypernyms (dog-animal) than to adjacent lower-order hypernyms (dog-poodle) but was unable to confirm the hypothesis. Understanding a word's superordinates (i.e., that the word is an instance of a broader category) may be a component of word knowledge that influences lexical processing and may explain variance in reading comprehension.

Definition knowledge. Unlike the other kinds of word knowledge described here, definitional knowledge involves both understanding something about a word and understanding something about a very unique academic genre. It is difficult to understand definitions, and children can easily misinterpret or misapply them. On the one hand, it has been amply demonstrated that definitions are hard to interpret, so providing children with a definition alone is not sufficient to ensure that they have an accurate representation of a word and how it is used (Miller & Gildea, 1987; Scott & Nagy, 1997). On the other hand, the combination of a definition with contextualized exposures to a word results in richer word learning (Bolger, Balass, Landen, & Perfetti, 2008; Gardner, 2007). For our purposes, the most relevant studies to date examined the extent to which additional variance in students' reading comprehension was explained by performance on a definition task, after controlling for their knowledge of the word's synonyms (Ouellette, 2006; Cain & Oakhill, 2014). These studies suggested that understanding a word's definitions explains additional variance in reading comprehension, although in these cases, latent scores were not used to model the relationships between these collinear predictors.

With some notable exceptions, the research on dimensions of word knowledge has been unable to parse how different aspects of word knowledge relate to students' reading comprehension outcomes. In this paper, we assess students on a common set of words across alternative assessment types to determine which dimensions of academic vocabulary knowledge

(i.e., multiword expressions, topical associates, hypernyms, definitions) explain variance in reading comprehension, controlling for academic vocabulary knowledge measured with a traditional synonym task.

Methods

Student Sample

The data for this study came from a three-year Institute for Education Sciences-funded randomized efficacy trial of Word Generation (www.wordgeneration.org). The current study uses data only from students who participated in the third year of the trial. The students were recruited from 12 middle schools from a large urban school district in California. The student sample for the current study consisted of 27% sixth graders, 37% seventh graders, and 36% eighth graders (Table 1). Approximately 10% of the sample was classified by the school district as limited English proficient (LEP). Sixty-three percent of the sample were eligible for the federally funded free and reduced lunch program.

<<<< Insert Table 1 >>>>

Measures

Reading comprehension. In order to assess students' reading skills, the passage comprehension subtest of the Gates-MacGinitie Reading Test was administered in the spring of the 2011–2012 school year. Level 6 Form T was administered to sixth-grade students, and level 7/9 Form T was given to seventh- and eighth-grade students (Table 1). Students were asked to read a passage and answer relevant comprehension questions (48 items). The extended scores were scaled such that a score of 516 corresponded to average achievement at the beginning of the

sixth grade and a score of 528 to average achievement at the beginning of the seventh grade. The reliability for our analytical sample was .92, and the mean and standard deviation were 530.60 and 39.59, respectively.

Vocabulary measures.

Target word sample. The words used in all our vocabulary measures were a subsample taken from Coxhead's (2000) Academic Word List, which are general academic words that are used across academic disciplines (Beck, McKeown, & Kucan, 2002). Figure 1 plots the frequency and dispersion estimates (Zeno, Ivens, Millard, & Duvvuri, 1995) of 2,000 words randomly selected from a much larger English corpus, and it overlays the target words from this study in darker markers. The target words in this study tended to be more dispersed across academic disciplines (M = .71, SD = .23) than words in the larger corpus (n = 150,000, M = .20, SD = .27). They also tended to be more frequently used (M = 45. 95, SD =6. 44) than typical words (M = 29.2, SD = 10.45). The findings reported here likely generalize to similar academic words but not to all words.

********** Figure 1 **********

Academic vocabulary test (Synonym task). Students' academic vocabulary knowledge was assessed with a 50-item multiple choice synonym test called the Academic Vocabulary Test, which was developed and validated by the research team (Snow, Lawrence & White, 2009; all test forms are uploaded in the IRIS digital depository). For each item, a target word was embedded in a short sentence, and students were asked to choose the closest synonym for the target word from four answer choices. Students' academic vocabulary knowledge was a covariate in our analysis. The reliability for our analytical sample was .92, and the mean and

standard deviation were 33.63 and 9.84, respectively. This test was administered in the fall of the 2011–2012 school year. Although this test is in a synonym task format, this test will be referred to as an academic vocabulary test throughout this discussion.

Depth measures. Four measures of vocabulary knowledge were developed and piloted by the research team at Educational Testing Service (i.e., multiple expressions, topical associates, hypernyms, and definitions; Deane et al., 2014). There were four forms of each type of assessment (4 forms x 4 test types = 16 forms of assessments), and each form consisted of 12 items. As this study was part of a randomized trial, students were also asked to take assessments and questionnaires that were not included in this current study; therefore, it was not possible to give each student the four test types, given time constraints and concerns about student test fatigue. We were also concerned that if students were asked to take all four types of depth measure, there would be testing effects. In other words, if students took the assessments on the same sets of target words that were presented in different ways, they might have been able to determine the answers through test exposure. Therefore, we created 16 small test forms and randomly distributed one to each student during their testing session for the larger study.

Multiword expressions. The multiword expressions assessment was designed to measure students' knowledge of the phrasal patterns that were characteristic of the target words. Figure 2 displays an example of an item in the multiword expressions measure. The stem of the item took the form of a cloze sentence-completion multiple choice item. Students were required to fill in the blank by selecting a word from three answer choices. All three answer choices belonged to the same part of speech. One of the distractors was a very bad choice—it was ungrammatical, distinctly odd-sounding, or awkward. The other distractor was less plausible than the key, not

widely attested, and not attractive, because it competed seriously with the key. The key was designed to be a natural, idiomatic, and relatively frequent collocate of the target word in the context of the sentence presented in the prompt. The average reliability for our analytical sample across four forms was .59, and the mean and standard deviation were 7.77 and 2.40, respectively.

********** Figure 2 **********

Topical associates. The topical-associates assessment was designed to assess whether students recognized associations between the target word and other words without requiring them to recognize a specific synonym or definition. It required that students understood words that were semantically associated with the target word, but it imposed no requirement that students understood exactly why or how the two words were related in meaning. In this test type, students were provided with three words (i.e., the stimulus) and were asked to select, from three answer choices, the word that was typically associated with these words. The three words that were presented in each item belonged to the same part of speech but not necessarily to the same part of speech as the target word. They were neither synonyms nor hypernyms of the target word. The relationship between the key and the stimulus reflected a true topical association and not a collocational pattern. In other words, it was expected that students might possess a clear sense of what the targeted word might be associated with without necessarily having specific knowledge of the word's meaning. An example of a topical associates item is presented in Figure 3. The average reliability for our analytical sample across four forms was .65, and the mean and standard deviation were 8.01 and 2.49, respectively.

********** Figure 3 **********

Hypernyms. The hypernym assessment was intended to measure whether students could recognize the broad meaning or category to which the target words belonged. In other words, this task required some knowledge of deeper semantic structures, although that knowledge did not need to be precise enough to provide definitions or identify synonyms. Students were required to identify the target word correctly through its membership in a superordinate class.

Each hypernym item was a cloze sentence-completion task with the target word in the item stem. The answer choices were words that could plausibly fill in the blank, belonged to the same part of speech, and were more or less at the same level of abstraction as the key. The two distractors were syntactically appropriate but semantically inappropriate and had a word frequency similar to that of the key. In this case, although students were expected to understand that the target word was a member of a broader category, they were not expected to be able to provide an exact definition of the word or generate semantically appropriate contexts in which the word could be used. The average reliability for our analytical sample across four forms was .64, and the mean and standard deviation were 7.26 and 2.55, respectively. An example of a hypernym item is presented in Figure 4.

********** Figure 4 **********

Definitions. The definitions assessment was intended to measure students' deepest degree of semantic knowledge of vocabulary by demonstrating their ability to distinguish between the definitions of closely related words. Students were presented with a target word and were asked to choose the correct definition from the answer options. The distractors were definitions of other academic target words.

This assessment deviated from the other assessment types in two very important ways.

First, four answer options were offered to the students instead of three—the item presented the target word in the stem. Second, the four answer options were designed to be equally attractive to a student who did not know the exact definition of the target word. As in the other three assessment types, the difficulty of the words represented by the answer options was controlled to be no higher than the difficulty of the target word that was being tested. The average reliability for our analytical sample across four forms was .54, and the mean and standard deviation were 5.48 and 2.37, respectively. An example of a definitions item is presented in Figure 5.

********** Figure 5 **********

Analytic Plan

******** Insert Figure 6 ********

Our principal interest in this study was to determine whether performance on the four types of vocabulary-depth measures (topical associates, multiword expressions, hypernyms, and definitions) could explain unique variance in reading comprehension skills, controlling for performance on an academic vocabulary test. In order to estimate the added influence of the depth measures, we used a residual factors approach (Bentler & Satorra, 2000) with latent variables (see Figure 6). To avoid attenuated associations between the variables resulting from measurement error, we created latent variables of all constructs by dividing each variable into two indicators, one for odd and one for even items. Correcting for measurement errors in this way can prevent and correct the potentially negative effects of measurement error on path and regression analysis (see Cole & Preacher, 2014). This approach was particularly useful, since we were concerned about the relatively low reliabilities of some of our assessments. In these models, it was only the common variance between the observed indicators that constituted the latent

construct (factor). The part of the observed indicators that contained the measurement errors was estimated (the residuals of the observed variables), but it was left out of the structural part of the model (i.e., the relations between the factors) where our hypotheses were being tested. All analyses were conducted with Mplus (Muthén & Muthén, 1998–2015), using full information maximum likelihood to handle missing values.

In these models, we regressed both vocabulary depth and reading comprehension on academic vocabulary (β1 and β2, respectively, in Table 6) as well as reading comprehension on the residual of vocabulary depth (β3 in Table 6). With this approach, β1s and β2s can be interpreted as ordinary bivariate regression coefficients. Squaring these standardized coefficients provides the percentage of variance in vocabulary depth and reading comprehension, respectively, that can be explained by academic vocabulary knowledge. The β3 is the standardized bivariate regression between reading comprehension and the part of vocabulary depth that is not common with performance on the synonym task. Squaring β3 gives the percentage of the variance in reading comprehension that vocabulary depth explains uniquely, that is, beyond the impact of an academic vocabulary test (synonym task). Adding the squares of β2 and β3 results in the amount of variance in reading comprehension that can be accounted for by academic vocabulary test performance and vocabulary depth (the R^2 of reading comprehension).

We estimated 20 such models, one for each of the 16 forms of depth measures and four using summary scores of the four forms in each of the categories (topical associates, multiword expressions, hypernyms, and definitions) of depth measures.

Results

****** Table 2. Correlations between item types, reading and synonym ******

****** Table 3. Difficulty of each item by form ******

Table 2 presents the correlations across each of the academic vocabulary measures and both academic vocabulary (synonym task) and the reading comprehension test. Not surprisingly, correlations across these measures were high. Given that we controlled for performance on the academic vocabulary synonym task in our analysis, we hypothesized that the measures with a relatively higher correlation with the synonym task (such as the hypernym task) would explain less unique variance in reading comprehension than the measures with a relatively lower correlation (such as the definition task). Table 3 presents the percentage of students who answered questions about each word correctly across forms and types.

****** Table 4. Overall performance for each assessment ******

Table 4 presents the mean scores for each vocabulary-depth assessment. The topical-associates task (M = 8.01, SD = 2.49) and multiword-expressions (M = 8.24, SD = 2.38) tasks were the easiest. Students found the definitions task most difficult (M = 5.48, SD = 2. 37). Table 5 presents the regression coefficients from our structural equation models (see Figure 6). The top four rows present the results from aggregate models for each form by type. That is, the first row presents results from a model fit to data from students who completed one of the four multiword-expressions depth measures. The second, third, and fourth rows present the model fit to data from students who completed one of the four topical-associates, hypernyms, or definitions assessments. These results demonstrate a strong relationship between the various

"depth" measures of students' vocabulary knowledge and their knowledge as measured by the 50-item synonym task (β1s ranging from .908 to .946, $p < .001$). A strong relationship between overall vocabulary knowledge and reading comprehension was also observed (β2s ranging from .851 to .894, $p < .001$). The results also demonstrate that the knowledge of multiword expressions, topical associates, and definitions (but not of hypernyms) explains unique variance in reading comprehension beyond overall vocabulary (β3s). The percentage of variance in reading comprehension that was uniquely explained by multiword expressions was 1.66% ($p < .001$), and it was 2.53% ($p < .001$) for topical associates and 8.24% ($p < .001$) for definitions.

The 16 lower rows of Table 5 separately present the results of the latent regression models for each form. Because the number of participants in these single-form models was approximately one-fourth the number of participants in the aggregated models, the power to detect significant relationships is considerably lower in the single-form models. As is evident, there is also variation between the forms within each category, and only five of the single forms suggest that the depth measures explain unique variance in reading comprehension beyond overall vocabulary. However, if we adjust for the number of different forms (16), only one of the single-form models (Form TA 4) shows a significant contribution of a depth measure to reading comprehension above overall vocabulary, indicating that with one exception, test forms were similar across measures in how well they explained differences in reading comprehension.

Discussion

The present study investigated the relationship between four novel tests of vocabulary and reading comprehension among middle school students. Even controlling for a 50-item

synonym vocabulary measure, some of these depth items explained unique variance in reading comprehension. Specifically, performance on the multiword-expressions task, the topical-associates task and the definitions task all explained unique variance in reading comprehension.

The literature review and the descriptive data provide a plausible explanation of why knowledge of multiword expressions might explain unique variance in reading comprehension beyond the synonym task. Our student participants included those with varying levels of English and from diverse linguistic backgrounds. Approximately 10% of the sample were limited English proficient (LEP) students. Multiword expressions are common in English writing, and the ability to comprehend them efficiently may facilitate comprehension above and beyond the knowledge of particular words in the expression. Native speakers of English are exposed to these expressions from a young age, and some of their earliest utterances are formulaic expressions (which they could not parse at the word level). Second-language learners, such as the LEP students in our sample, acquire knowledge of these expressions slowly and may process them differently (Ellis et al., 2008). Descriptive data demonstrated that for LEP students, this was a more difficult item type. Taken together, these results suggest that the multiword expressions task measures the kind of vocabulary knowledge that differentiates skill at the student level and that is leveraged in reading comprehension.

The topical-associates task was also highly correlated with the reading comprehension measure and explained unique variance in reading, even controlling for performance on the 50-item synonym task. One of the challenges of this task is sorting between the multiple meanings of words used in the task to establish a coherent pattern of relationships across the words. Because the synonym task presents a sentence and a synonym, even if a target word has multiple meanings, the alternative meanings are unlikely to interfere with or support students' completion

of such a highly semantically contextualized task. The topical-associates task may assess knowledge connected to how many meanings a student knows and how well they can navigate multiple meanings of words simultaneously to find connections across them. We intend to test this hypothesis by examining whether this task is more difficult for students when the target words are polysemous (controlling for frequency and orthographic, phonological, and semantic features).

The definitions task had the strongest relationship with reading comprehension after controlling for the synonym task ($\beta = .287^{***}$, $p < 0.001$). This finding is a replication of Ouellette (2006), who also found that student knowledge of definitions explained unique variance in reading, with controls for student vocabulary breadth. It could be that the ability to complete the synonym task is a prerequisite to the ability to perform well on a definitions assessment using similar words. That is, in Ouellette's terms, the definitions task measures the same kind of knowledge as the synonym task but at a deeper level. On the other hand, it may be that the definitions assessment relies on verbal metacognition more heavily than does the synonym task. In other words, for students to complete the definitions task, they had to read and comprehend four definitions, and it is the ability to read this academic genre that accounts for why this measure explains additional variance in reading.

The only item type that did not predict additional variance in reading comprehension, controlling for the academic vocabulary test, was the hypernym measure. Again, it is possible that the idea of "depth" explains the relationship between the synonym and hypernym tasks. At least with respect to vocabulary knowledge as leveraged for reading, the ability to identify a synonym might be "deeper" than the kind of knowledge used to complete the hypernym task.

These analyses suggest that even with controls for synonym knowledge of the same set of

academic words, the new alternative assessments of vocabulary knowledge explain additional variance in reading. Knowing how a word is typically used in a multiword expression may not have a straightforward relationship with knowledge of definitions or synonyms; rather, although different ways of knowing may be related (as per the aptitude hypothesis), individual and group level trends across measurement dimensions may preclude simple unidimensional characterizations. There is no evidence for a linear conceptualization of "vocabulary depth" across these measures in terms of how they related to reading.

These results help us better understand the hypothesized relationships between vocabulary and reading summarized above as the instrumentalist, world knowledge, aptitude, and lexical access hypotheses. Given that the sets of words in each of the researcher-developed measures were the same, these results provide definite evidence that the instrumentalist hypothesis does not completely explain the relationship between reading and vocabulary. This aligns with the limited purview afforded this hypothesis in most of the research literature. These tasks could be considered demonstrations of world knowledge. For instance, being able to identify definitions could be considered a demonstration of knowledge of an academic register, so these results might be interpreted as supporting a version of the knowledge hypothesis (i.e., the strong relationship of vocabulary knowledge and reading ability is explained by the reliance of each on world knowledge). Note, however, that this version of the knowledge hypothesis extends beyond the idea that knowledge of a word is knowledge of the world to the idea that the way that we know a word is knowledge of the world. All measures were highly correlated, as would be predicted by the aptitude hypothesis, but we do not have lexical access data to test the variance explained by lexical retrieval efficiency.

It might be argued that our conceptualization of "vocabulary knowledge" is too

expansive; it might be suggested that we are controlling for vocabulary knowledge with the synonym tasks, but the novel measures are not assessments of vocabulary *per se*. If we push the above suggested argument further, we might argue that the definitions task is not a vocabulary task at all but is a test of academic register awareness. Similarly, the multiword expressions task might be conceptualized as assessing the knowledge of connectives and discourse markers (Uccelli, Galloway, & Barr, 2015). From our point of view, these are all aspects of vocabulary knowledge. We understand vocabulary as an expansive domain with biological (Huth, de Heer, Griffiths, Theunissen & Gallant, 2016), affective (Warriner, Kuperman, & Brysbaert, 2013), cognitive (Perfetti, & Stafura, 2016), and historical (Vanhove, 2008) dimensions. We invoke an expansive view of vocabulary knowledge informed by all these dimensions to account for the strong relationship between vocabulary and comprehension (especially when we use latent scores derived from normed measures) and the results described here.

There are important limitations to this study. For example, we did not have enough students to fit multigroup models across all the measures simultaneously. Nor were we able to run multigroup comparisons according to students' language status (LEP vs. non-LEP). Our test reliabilities were not as high as we would like, partly because we only had 12 items for each form of new vocabulary measure. Yet another limitation is that we could not compare the four depth measures in the same model. That is, we do not know whether it is the shared or unique variance of these measures that explain reading comprehension beyond the academic vocabulary measure (synonym task). Because we only use one reading comprehension measure, the impact of the various depth measures may vary as a function of this. Nonetheless, this study provides a valuable contribution in showing that words are known in multiple ways and that understanding different components of vocabulary knowledge can help us model student reading performance

more accurately.

References

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles—A problem of polysemy. *Cognitive Psychology*, *7*(2), 167–180.

Anderson, R. C., Pichert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. V., & Trollip, S. R. (1976). Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior*, *15*(6), 667–679. doi:10.1016/0022-5371(76)90059-1

Beck, I., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford.

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

Bentler, P. M., & Satorra, A. (2000). Hierarchical regression without phantom factors. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(2), 287–291. doi:10.1207/S15328007SEM0702_8

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405.

Bolger, D., Balass, M., Landen, E., & Perfetti, C. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, *45*(2), 122–159.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford University Press.

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension. *L'Année Psychologique*, *114*, 647–662.

Carroll, J. B. (1941). A factor analysis of verbal abilities. *Psychometrika*, *6*(5), 279–307. doi:10.1007/BF02288585

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Memory and Language*, *8*(2), 240–247.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89. doi:10.1093/applin/amm022

Coxhead, A. (2000). A new academic word list. *Tesol Quarterly*, *34*(2), 213–238.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, 99(2), 311-325. doi:10.1037/0022-0663.99.2.311

Deane, P., Lawless, R. R., Li, C., Sabatini, J., Bejar, I. I., & O'Reilly, T. (2014). Creating vocabulary item types that measure students' eepth of semantic knowledge. *ETS Research Report Series*, *2014*(1), 1–19. doi:10.1002/ets2.12001

Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: the current state of the literature. *Psychonomic Bulletin & Review*, *22*(1), 13–37. https://doi.org/10.3758/s13423-014-0665-7

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1–44.

doi:10.1080/19345740802539200

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375–396. doi:10.1002/j.1545-7249.2008.tb00137.x

Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language*, *67*(4), 407–425. doi: 10.1016/j.jml.2012.07.010

Gardner, D. (2007). Children's immediate understanding of vocabulary: Contexts and dictionary definitions. *Reading Psychology*, *28*(4), 331–373.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10. doi: 10.1177/074193258600700104

Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: acquiring depth of knowledge through network building. *Vigo International Journal of Applied Linguistics*, *10*(2), 221–240. doi:10.1111/j.1473-4192.2000.tb00149.x

Halff, H. M., Ortony, A., & Anderson, R. C. (1976). A context-sensitive representation of word meanings. *Memory & Cognition*, *4*(4), 378–383. doi:10.3758/BF03213193

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Hyland, K. (2012). Bundles in Academic Discourse. *Annual Review of Applied Linguistics*, *32*, 150–169. doi:10.1017/s0267190512000037

Jenkins, J. J. (1970). The 1952 Minnesota word association norms. In L. Postman & G. Keppel (Eds.), *Norms of Word Association* (pp. 1–38). London, UK: Academic Press.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(3), 802–823.doi:10.1037/a0030859

LaBerge, D., & Samuels, S. J. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. doi: 10.1016/0010-0285(74)90015-2

Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word Generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal*, *52*(4), 750–786. doi.org/10.3102/0002831215579485

Lawrence, J. F., Francis, D., Paré-Blagoev, J., & Snow, C. E. (2017). The Poor Get Richer: Heterogeneity in the Efficacy of a School-Level Intervention for Academic Language. *Journal of Research on Educational Effectiveness*, *10*(4), 767–793. doi.org/10.1080/19345747.2016.1237596

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of Academic Vocabulary Instruction for Linguistically Diverse Adolescents: Evidence From a Randomized Field Trial. *American Educational Research Journal*, *51*(6), 1159–1194. doi.org:10.3102/0002831214532165

Lesaux, N. K., Kieffer, M. J., Faller, E., & Kelley, J. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly*, *45*(2), 196–228.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie reading test technical report: Forms S and T*. Chicago, IL: The Riverside Publishing

Company.

Manning, C. D., & Schuetze, H. (1999). *Foundations of statistical natural language processing, collocations*. Cambridge, MA: The MIT Press.

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, *45*(2), 267–290. doi:10.5054/tq.2011.247708

McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior*, *15*(1), 3–18.

Miller, G. A., & Gildea, P. M. (1987). How children learn words. *Scientific American*, *257*(3), 94–99.

Nagy, W. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52–77). New York, NY: Guilford Press.

Oakhill, J., Cain, K., McCarthy, D., & Nightingale, Z. (2012). Making the link between vocabulary knowledge and comprehension skill. In A. Britt, S. Goldman & J.-F. Rouet (Eds.), *From Words to Reading for Understanding* (pp. 101–114). New York: Routledge

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*(3), 554–566. doi:10.1037/0022-06

Pany, D., Jenkins, J. R., & Schreck, J. (1982). Vocabulary instruction: Effects on word knowledge and reading comprehension. *Learning Disability Quarterly*, *5*(3), 202–215. doi:10.2307/1510288

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of*

Reading, *11*(4), 357–383. doi:10.1080/10888430701530730

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

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. doi:10.1080/10888438.2013.827687

Perfetti, C., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*, 1281–1292.

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: a latent change score modeling study. *Child Development*, *86*(1), 159–175. doi:10.1111/cdev.12292

Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, *17*(6), 415–434.

Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, *30*(2), 178.

Roediger, H. L. I., & McDermott, K. B. (1999). False alarms and false memories. *Psychological Review*, *106*(2), 406–410. doi:10.1037/0033-295X.106.2.406

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(4), 803–814. doi:10.1037/0278-7393.21.4.803

Schmitt, N., Jiang, X., & Grabe, W. (2011). The Percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26–43. doi:10.1111/j.1540-4781.2011.01146.x

Scott, J. A., & Nagy, W. E. (1997). Understanding the cefinitions of unfamiliar verbs. *Reading Research Quarterly*, *32*(2), 184–200. doi:10.1598/RRQ.32.2.4

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, *2*(4), 325–344. doi.org/10.1080/19345740903167042

Snow, C. E., Porche, M. V., Tabors, P., & Harris, S. R. (2007). *Is literacy enough?: Pathways to academic success for adolescents*. Baltimore, MD: Paul H. Brookes Publishing Co.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292.

Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings*. Mahwah, NJ: Lawrence Erlbaum.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, *56*(1), 72–110.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407.

Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *The American Psychologist*, *38*(8), 878–893.

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), 381–398.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology, 45*(4), 1097-1118.

Tunmer, W. E., & Herriman, M. L. (1984). The development of metalinguistic awareness: A conceptual overview. In W. E. Tunmer, C. Pratt, & M. L. Herriman (Eds.), *Metalinguistic Awareness in Children* (pp. 12–35). Berlin: Springer-Verlag. doi:10.1007/978-3-642-69113-3_2

Uccelli, P., Galloway, E. P., & Barr, C. D. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly, 50*(3), 337-356. doi:10.1002/rrq.104

Vanhove, M. (2008). *From polysemy to semantic change: towards a typology of lexical semantic associations* (Vol. 106). Amsterdam, Netherlands: John Benjamins Publishing.

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*(1), 8–25. doi:10.1080/10888438.2011.536125

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(02), 217–234.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., … Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: a 5-year longitudinal study. *Developmental Psychology*, *33*(3). 469-479.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. doi:10.3758/s13428-012-0314-x

Weekes, B. S., Hamilton, S., Oakhill, J. V., & Holliday, R. E. (2008). False recollection in children with reading comprehension difficulties. *Cognition*, *106*(1), 222–233. doi:10.1016/j.cognition.2007.01.005

Wilkins, A. J. (1971). Conjoint frequency, category size, and categorization time. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 382–385. doi:10.1016/S0022-5371(71)80036-1

Wittgenstein, L. (1958). *Philosophical investigations*. (G. E. M. Anscombe, Trans.) (2nd ed.). Oxford, UK: Basil Blackwell.

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, *18*(4), 742–750.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

Table 1. *Assessment type for participating students by grade level*

Test Type	N	Grade 6	Grade 7	Grade 8
Multiword Expressions	1,478	19%	35%	46%
Hypernym	1,187	21%	48%	31%
Topical	1,395	23%	39%	38%
Definitions	1,416	48%	25%	27%
Total	5,855	27%	37%	36%

Table 2. *Correlations between student total performance on each item type, academic vocabulary, and reading comprehension*

Item Type	Academic Vocabulary	Reading Comprehension
Multiword Expressions (*n* = 1,479)	0.67	0.66
Topical Associates (*n* = 1,364)	0.69	0.70
Hypernyms (*n* = 1,285)	0.70	0.69
Definitions (*n* = 1,470)	0.65	0.67

Table 3. *Average student performance on each target word by type by form*

	Target Word	Multiword Expressions		Topical Associates		Hypernyms		Definitions	
		Form E		Form A		Form I		Form M	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	Alter	0.91	*0.29*	0.76	*0.43*	0.60	*0.49*	0.51	*0.50*
2	Controversy	0.49	*0.50*	0.78	*0.41*	0.69	*0.47*	0.38	*0.49*
3	Component	0.59	*0.49*	0.60	*0.49*	0.68	*0.47*	0.50	*0.50*
4	Generate	0.72	*0.45*	0.67	*0.47*	0.67	*0.47*	0.43	*0.50*
5	Allocated	0.68	*0.47*	0.54	*0.50*	0.43	*0.50*	0.35	*0.48*
6	Emphasize	0.48	*0.50*	0.82	*0.39*	0.62	*0.49*	0.30	*0.46*
7	Cohesion	0.79	*0.40*	0.49	*0.50*	0.33	*0.47*	0.32	*0.47*
8	Assess	0.44	*0.50*	0.56	*0.50*	0.31	*0.46*	0.32	*0.47*
9	Recite	0.74	*0.44*	0.82	*0.39*	0.78	*0.42*	0.69	*0.46*
10	Eliminated	0.83	*0.38*	0.74	*0.44*	0.89	*0.31*	0.90	*0.30*
11	Enforced	0.55	*0.50*	0.85	*0.36*	0.53	*0.50*	0.59	*0.49*
12	Amnesty	0.33	*0.47*	0.48	*0.50*	0.30	*0.46*	0.40	*0.49*
		Form F		Form B		Form J		Form N	
13	Distinct	0.43	*0.50*	0.87	*0.34*	0.53	*0.50*	0.43	*0.50*
14	Invasion	0.82	*0.39*	0.89	*0.31*	0.87	*0.33*	0.58	*0.49*
15	Relevance	0.70	*0.46*	0.64	*0.48*	0.58	*0.49*	0.56	*0.50*
16	Disproportionately	0.39	*0.49*	0.54	*0.50*	0.58	*0.49*	0.45	*0.50*
17	Assume	0.63	*0.48*	0.71	*0.45*	0.45	*0.50*	0.61	*0.49*
18	Retain	0.54	*0.50*	0.55	*0.50*	0.58	*0.49*	0.48	*0.50*
19	Constrain	0.76	*0.42*	0.63	*0.48*	0.60	*0.49*	0.31	*0.46*
20	Enable	0.77	*0.42*	0.58	*0.49*	0.60	*0.49*	0.65	*0.48*
21	Document	0.21	*0.41*	0.86	*0.34*	0.73	*0.44*	0.39	*0.49*
22	Perceive	0.28	*0.45*	0.47	*0.50*	0.45	*0.50*	0.30	*0.46*
23	Subsequent	0.50	*0.50*	0.68	*0.47*	0.58	*0.49*	0.37	*0.48*
24	Incentives	0.59	*0.49*	0.70	*0.46*	0.44	*0.50*	0.20	*0.40*
		Form G		Form C		Form K		Form O	
25	Contaminate	0.84	*0.37*	0.82	*0.39*	0.59	*0.49*	0.57	*0.50*
26	Prescribe	0.90	*0.29*	0.76	*0.43*	0.66	*0.47*	0.07	*0.26*
27	Intrinsic	0.67	*0.47*	0.37	*0.48*	0.40	*0.49*	0.25	*0.43*
28	Attribute	0.51	*0.50*	0.76	*0.43*	0.45	*0.50*	0.40	*0.49*
29	Outweigh	0.69	*0.46*	0.75	*0.43*	0.52	*0.50*	0.57	*0.50*
30	Amend	0.25	*0.43*	0.52	*0.50*	0.59	*0.49*	0.53	*0.50*
31	Obtain	0.95	*0.22*	0.76	*0.43*	0.78	*0.41*	0.50	*0.50*
32	Comprise	0.43	*0.50*	0.63	*0.48*	0.40	*0.49*	0.36	*0.48*
33	Eligible	0.82	*0.38*	0.83	*0.37*	0.80	*0.40*	0.56	*0.50*
34	Extracted	0.70	*0.46*	0.63	*0.48*	0.85	*0.36*	0.50	*0.50*
35	Conserve	0.73	*0.44*	0.83	*0.38*	0.84	*0.37*	0.58	*0.49*
36	Complex	0.82	*0.38*	0.81	*0.39*	0.58	*0.49*	0.42	*0.49*
		Form H		Form D		Form L		Form P	
37	Restrict	0.97	*0.18*	0.58	*0.49*	0.80	*0.40*	0.32	*0.47*
38	Attained	0.73	*0.45*	0.66	*0.48*	0.75	*0.44*	0.48	*0.50*
39	Exclude	0.85	*0.36*	0.44	*0.50*	0.79	*0.41*	0.71	*0.45*
40	Aptitude	0.65	*0.48*	0.48	*0.50*	0.53	*0.50*	0.36	*0.48*
41	Equity	0.82	*0.38*	0.65	*0.48*	0.64	*0.48*	0.26	*0.44*
42	Critical	0.70	*0.46*	0.54	*0.50*	0.67	*0.47*	0.50	*0.50*
43	Contrast	0.40	*0.49*	0.78	*0.42*	0.70	*0.46*	0.65	*0.48*
44	Interaction	0.88	*0.32*	0.72	*0.45*	0.68	*0.47*	0.63	*0.48*
45	Compatible	0.60	*0.49*	0.79	*0.41*	0.72	*0.45*	0.59	*0.49*
46	Acquired	0.44	*0.50*	0.52	*0.50*	0.67	*0.47*	0.40	*0.49*
47	Paralyzed	0.83	*0.38*	0.82	*0.38*	0.87	*0.34*	0.85	*0.35*
48	Apathy	0.71	*0.46*	0.68	*0.47*	0.35	*0.48*	0.28	*0.45*

Table 4. *Overall performance on each item type*

	Total Sample	
	M	*SD*
Multiword Expressions	7.77	2.40
Topical Associates	8.01	2.49
Hypernyms	7.26	2.55
Definitions	5.48	2.37

Table 5. *Regression coefficients table from structural equation modeling*

	Academic Vocabulary -> Depth		Academic Vocabulary -> Reading Comprehension		Depth > Reading Comprehension		
	β1	95% CI	β2	95% CI	β3	95% CI	R^2
Multiword expressions	.908***	.865-.951	.894***	.876-.911	.129***	.049-.209	.815***
Topical associates	.899***	.864-.934	.890***	.872-.907	.159***	.090-.229	.817***
Hypernyms	.915***	.878-.953	.877***	.846-.896	.075	-.004-.154	.775***
Definitions	.946***	.894-.998	.851***	.829-.873	.287***	.124-.450	.807***
Form MW 1	.949***	.870-1.029	.874***	.832-.915	.258	-.009-.524	.830***
Form MW 2	.958***	.803-1.112	.875***	.833-.917	.214	-.257-.684	.811***
Form MW 3	.963***	.902-1.024	.898***	.865-.930	.072	-.121-.266	.811***
Form MW 4	.972***	.905-1.039	.915***	.886-.943	.23	-.075-.536	.889***
Form TA 1	.917***	.847-.987	.928***	.898-.959	.037	-.104-.177	.863***
Form TA 2	.931***	.854-1.009	.907***	.874-.941	.313**	.104-.521	.921***
Form TA 3	.904***	.191-.250	.879***	.887-1.076	.096	-.172-.1.099	.783***
Form TA 4	.872***	.798-.946	.851***	.811-.891	.236***	.098-.373	.780***
Form H 1	.875***	.792-.958	.883***	.846-.920	.145*	.012-.278	.801***
Form H 2	.886***	.811-.960	.881***	.847-.915	.185**	.054-.316	.810***
Form H 3	.990***	.922-1.057	.853***	.813-.893	.07	-.438-.577	.733***
Form H 4	.961***	.900-1.022	.896***	.856-.937	.128	-.145-.401	.820***
Form D 1	.917***	.815-1.019	.791***	.793-.842	.201	-.021-.427	.665***
Form D 2	.913***	.780-1.045	.836***	.786-.886	.18	-.067-.427	.731***
Form D 3	.979***	.896-1.061	.871***	.832-.911	.422	-.378-1.222	.937**
Form D 4	.906***	.817-.995	.899***	.862-.935	.248**	.072-.423	.869***

Note. *** $p < .001$, ** $p < .01$, MW = Multiword expressions, TA = Topical associates, H = Hypernyms, D = Definitions

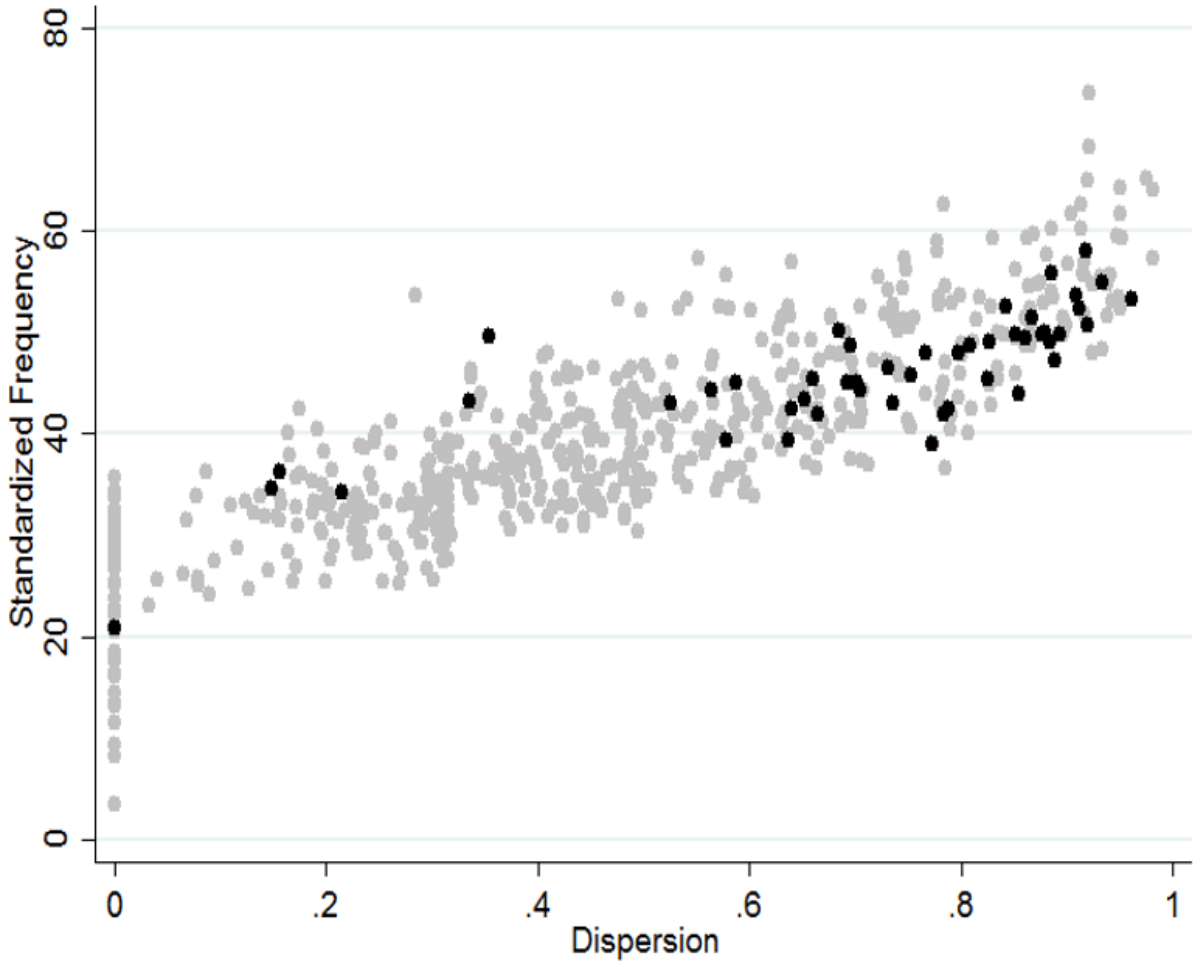

Figure 1. *Plot of frequency and dispersion estimates of 2,000randomly chosen words from English corpus. Target words selected for this study are displayed in darker markers.*

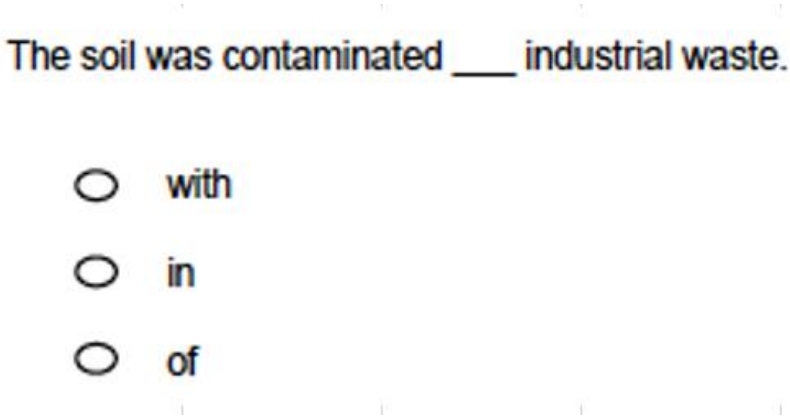

Figure 2. *A sample item in the multiword expressions measure.*

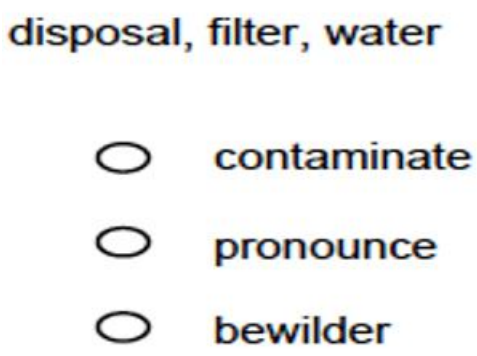

Figure 3. *A sample item in the topical associates measure.*

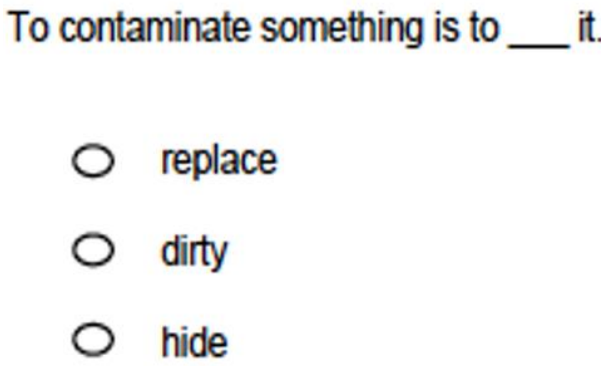

Figure 4. *A sample item in the hypernyms measure.*

Figure 5. *A sample item in the definitions measure*.

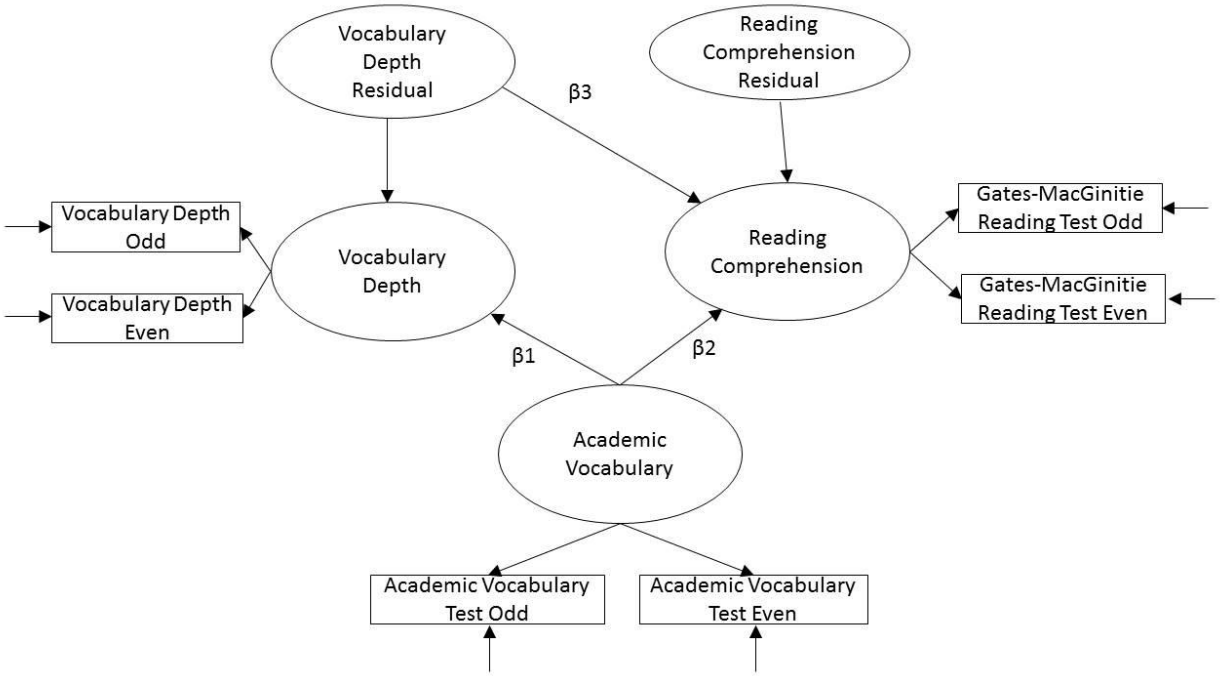

Figure 6. *The analytical mode that was tested in the current study. One model for each of the 16 forms of the depth measures and four models using summary scores of the four forms in each category were estimated.*