CrossMark

Using new vocabulary in writing: Exploring how word and learner characteristics relate to the likelihood that writers use newly taught vocabulary

Christina L. Dobbs[1] · Devin Kearns[2]

Abstract Understanding academic vocabulary is essential to student success in school. Use of academic vocabulary words in writing is considered one of the strongest measures of how well a reader understands a given word. In theory, willingness to use academic vocabulary in writing indicates the complexity of acquiring representations of the word's orthography, phonology, and semantics based on the word's characteristics. In addition, a learner's overall literacy skills should relate to whether they attempt to use words. In the present study, sixth graders ($n = 66$), seventh graders ($n = 60$), and eighth graders ($n = 41$) learned 25 academic vocabulary words in a supplementary academic vocabulary intervention and were then asked to use those words in short persuasive essay drafts. We measured whether these students attempted to use a word (a binary *uses* outcome) and the number of times a student used a word (a continuous *attempts* outcome) and used cross-classified random effects models to examine how (a) the orthographic, phonological, and semantic characteristics of words and (b) the students' literacy-related characteristics related to their uses and attempts. For word characteristics, students were more likely to use and attempt high frequency than low frequency words. For student characteristics, students proficient on the state exam were more likely to use and attempt the words, and students learning English were less likely to attempt the words. Implications for vocabulary intervention and writing instruction are discussed.

Keywords Vocabulary · Academic vocabulary · Writing

✉ Christina L. Dobbs
   cdobbs@bu.edu

[1] School of Education, Boston University, 2 Silber Way, Boston, MA 02215, USA

[2] Department of Educational Psychology, Neag School of Education, University of Connecticut, Storrs, CT, USA

✑ Springer

Introduction

In the transition from elementary to secondary school, adolescents should experience tremendous growth in the use of academic language (Nagy & Townsend, 2012). The volume of academic vocabulary in texts increases (Biemiller, 1999), and adolescents' teachers are more likely to use the complex forms that characterize academic vocabulary. Consequently, success in content-area learning requires that students add these words to their receptive vocabularies—and ultimately their expressive ones. Further, students must use these words in writing across disciplines, not only in speaking. Whether students use words in their writing is also important as an index of student ownership of new words (McKeown, Beck, & Sandora, 2012), as students can recognize and understand more words than they use in writing, likely because writing requires greater understanding of words than reading (Durso & Shore, 1991; Wesche & Paribakht, 1996). Unfortunately, expressive use of academic vocabulary is rarely required of students, either in classroom instruction (Beck, McKeown, & Kucan, 2002; Gersten, Dimino, Jayanthi, Kim, & Santoro, 2010; Nagy & Townsend, 2012) or assessment (Pearson, Hiebert, & Kamil, 2012).

Given the potential importance of written vocabulary use as an index of vocabulary knowledge and a contributor to writing quality generally, this paper is concerned with the factors that contribute to students' use of academic vocabulary in their writing. We want to understand what word and learner characteristics explain whether students attempt to use academic vocabulary. Our rationale is that understanding the word and child characteristics related to writing words could provide evidence to guide research on how to promote productive use of academic language.

Writing, the lexical quality hypothesis and word characteristics

This study takes a cognitive process perspective on the writing process, using Flower and Hayes' (1981) model. This model views writing as the orchestration of a variety of processes during composition related to an overall goal for a text. In this model, the translation of ideas into text is one part of a more complex process that also includes planning, reviewing, and monitoring processes that must be simultaneously employed. Translation is a process that involves filtering the task demands and expectations down into particular syntactic and lexical choices, a process that can be challenging for short-term memory.

Our work is grounded in the lexical quality hypothesis (LQH) because our hypothesis is that the likelihood a learner will use a word in their writing derives from the quality of the learner's representation for that word. Perfetti (2007) defined lexical quality as "the extent to which a mental representation of a word specifies its form and meaning components in a way that is both precise and flexible" (p. 359) and hypothesized this lexical quality as an element of key consequence for reading skills, including comprehension of text. Put differently, the LQH suggests that

whether learners acquire representations of words depends on their knowledge of the orthographic, phonological, and semantic dimensions of words (Perfetti, 2007; Perfetti & Hart, 2001, 2002). The LQH has been shown empirically to predict the development of reading comprehension in a variety of contexts including over time (Andrews & Bond, 2009; Goodwin, Gilbert, & Cho, 2013; Hersch & Andrews, 2012; Richter, Isberner, Naumann, & Neeb, 2013; Verhoeven & van Leeuwe, 2008; Verhoeven, van Leeuwe, & Vermeer, 2011). The hypothesis here is that knowledge that contains deep and robust representations of words might make those words easier to use in written text, as writers are unlikely to try to use words they do not know the meanings of when generating text (Baumann, Kame'enui, & Ash, 2003), especially in the cognitively complex context of generating academic text by novice writers. They are also unlikely to try to use words they can not spell or do not know how to read when generating text (Perfetti & Hart, 2001, 2002), meaning that uses of words in written text reflects some basis of word knowledge.

It is also clear that the orthographic, phonological, and semantic aspects of word knowledge develop at different rates, depending on the difficulty the word poses along a given dimension (e.g., Kearns & Al Ghanem, 2014). For example, the orthographic and phonological quality for the word *function* grows quickly because it is relatively easy to learn to spell and pronounce. *Function* has common letter-sound patterns and a common suffix (-*ion*). We expect, therefore, that the frequency of a word's orthographic patterns—represented here by *bigram frequency*—will facilitate students' ability to use the word in writing as they should be able to learn its spelling more quickly. To wit, Perfetti (1992) pointed out that the ability to spell a word provides evidence of a strong lexical representation. Kearns (2015) observed a facilitative effect of bigram frequency in word naming for third and fourth grade students, so we think this could translate to writing. Similarly, it should be easier to acquire phonological representations for words in denser phonological neighborhoods than sparse ones as the words contain familiar patterns. We test this possibility using a *phonological neighborhood size* metric with sensitivity for lengthy words.

In writing, however, we anticipate that variables that relate to semantics, including those related to morphology, would have effects as well. The quality of the semantic representation for *function,* for example, may grow slowly because the word is abstract and lacks a simple *imageable* definition. A key characteristic of academic vocabulary words, such as *function,* is that they are abstract in nature, which can pose a challenge for retaining a definition (Nagy & Townsend, 2012). The *surface frequency* of a word may affect the likelihood of use. Developing writers are more likely to have acquired strong conceptual understandings of high frequency words, such that they are more comfortable using them. The *size of a word's morphological family* may have similar effect: Individuals acquire better representations of words when they are part of morphological families that contain more entries (e.g., Ford, Davis, & Marslen-Wilson, 2010). For example, *rely* has no morphological neighbors, forcing writers to depend solely on the specific word's semantics to acquire knowledge about it. So, in practical terms, a student might be able to spell and pronounce *function* before understanding its meaning and perhaps long before using it herself/himself.

In summary, words' characteristics may affect whether students attempt to write words spontaneously, even when they are encouraged to do so. In a lexical quality framework, orthographic, phonological, and semantic and morphological characteristics may have independent influences of the likelihood of use in writing. A key focus of this paper is to investigate which characteristics influence the likelihood of use.

The data from Goodwin et al. (2013) provide evidence for this phenomenon. A reader's knowledge of the orthographic, phonological, and semantic characteristics of a word's root related to an adolescent's ability to spell, read, and indicate familiarity with a particular word. In turn, an adolescent's ability to do these things related to their reading comprehension, morphological awareness, and vocabulary. Put differently, a word's root characteristics relate to the quality of adolescents' representations individual words, and that individual word representation quality relates to adolescents' broader literacy skills. The Goodwin et al. data, therefore, indicate the importance of examining word and root characteristics that affect learners' use of these words in their writing.

Prior research also demonstrates the importance of relationships between lexical quality and word use in writing, though there remains much to explore. The use of more complex lexical items has a positive relationship with writing being perceived as higher quality at a variety of age levels and in several languages (Kim, Park, & Park, 2013; Maylath, 1996; Neilsen & Piché, 1981). There has been demonstration that the diversity of vocabulary used by a writer and the use of less frequent vocabulary are influenced by development (Olinghouse & Leaird, 2009). Additionally, there have been a few attempts to measure the use of complex vocabulary and rare words in writing, including one that attempted successfully to use rare words as a measure of expressive vocabulary (Flinspach, Scott, & Vevea, 2010) and a small set that measure the use of newly taught words in writing (see Duin & Graves, 1987; Lee, 2003; Lee & Muncie, 2006; Mancilla-Martinez, 2010 for examples).

We contend that exploration of word characteristics related to lexical quality is important to understand what affects the likelihood learners use academic vocabulary in their expressive language. Writing skill has a strong relation with performance in secondary and post-secondary education and in the workplace (National Commission on Writing, 2004, 2005), and difficulty using academic vocabulary increases the cognitive demand of writing—which is already very cognitively-demanding (Bonin, Fayol, & Gombert, 1997; Fayol, 2012). Despite the importance of both writing and vocabulary to student success, there is still much to learn about expressive uses of vocabulary, because knowing a word often requires a high level of knowledge of that work, such that not knowing a word is more likely to result in the writer not using the word (Baumann, Kame'enui, & Ash, 2003). To date, no study has considered how words' orthographic, phonological, and semantic characteristics might impact learners' willingness to use these words. Examining these characteristics is an important focus of the study.

Effects of learner characteristics

Examining how academic words' characteristics will affect whether learners try to write them is important and novel, but it is only part of the puzzle. Whether learners develop high-quality representations of words' three dimensions also relates to their literacy skills. For example, Perfetti and Hart (2002) presented data that a learner's ability to acquire a representation for an unfamiliar low frequency word depended on whether the learner was a stronger or weaker reader.

Much more could be known about how characteristics of writers contribute to the development of lexical representations, but there is evidence that individual differences in readers' skills and motivation play an important role in the development of reading comprehension over the elementary and subsequent years (Klauda & Guthrie, 2015; Ritchey, Silverman, Schatschneider, & Speece, 2015; Wigfield & Eccles, 2000). Readers who have strong comprehension are more likely to read than those with weaker skills, which results in the ability to learn more words over time (Stanovich, 1986). This increased word learning is associated with higher quality word representations because strong readers with large vocabularies have more representations already upon which to build new representations (Nagy, Anderson, & Herman, 1987; Perfetti, 2007). Thus, there is reason to expect that students with better reading skills develop better representations of words—orthographic, phonological, and semantic and morphological—and thus may attempt to use new vocabulary more frequently in writing. As described in the section on word characteristics, we anticipate that all features of representations should relate to the likelihood of word use. This is reflected in broad reading skill: Better readers have better overall lexical quality and thus may have the skill to write words better than their poor-reading peers (Perfetti, 2007). Whether having better reading skills is also related to learners' attempts to use words in text is as yet unclear, although theory suggests that students with better reading have better writing skills and should use academic words more often when they write (Juel, 1988).

This study is also concerned with particular ways that students from different language backgrounds use academic language in writing (Corson, 2002). We expect—in line with the proposal just described—that English learners (ELs) are less likely to use words than their non-EL peers. Even after students receive instruction on these new words—as all students in this study did—those who are still learning English may not have been able to acquire high quality representations and thus would not be likely to attempt the use of a word in writing.

The current study

To summarize, whether adolescents use academic vocabulary in their writing depends on the quality of their lexical representations for these words along orthographic, phonological, and semantic dimensions. It is suggested that the quality of these representations is partly governed by how easy it is to spell, say, and

indicate the meaning of these words—the words' characteristics. Further, these characteristics will affect the likelihood a learner will attempt to use a word in writing. Finally, the likelihood a word is used in writing relates to the learner's reading skills and whether the student is an EL, such that students with better reading skills and who are native English speakers should be more likely to try to use words in their writing.

This study, therefore, concerns middle schoolers' use of 25 newly-taught academic vocabulary words in short persuasive essay drafts within the context of a supplementary academic vocabulary intervention. We explore word characteristics and student characteristics that could be associated with the building of lexical representations that relate to students' written uses of the words.

This study is guided by two related research questions. Both questions concern word uses in writing during writing exercises in which students were encouraged— but not required—to use the words. The questions consider the characteristics of words students were taught that might affect whether they are used in writing and whether student characteristics affect word use. The difference in the questions relates to the dependent variable used. For the first research question, a binary outcome was considered, whether a reader ever attempted the use of a word in the writing exercises—regardless of the number of times they did so, an outcome we term an *attempt*. The second question uses a continuous outcome, namely the number of times (termed *uses*) that a student used a word in the writing exercises.

The specific questions are these: First, what word characteristics and student abilities relate to the likelihood a student would attempt to use a word in writing assignments designed to encourage word use? Second, what word characteristics and student abilities relate to the number of times a student used a word in the targeted writing exercises? As described above—and discussed in detail below— word characteristics were selected to be related to dimensions of lexical quality, bigram frequency (orthographic), phonological neighborhood size (phonological), imageability and morphological family size (semantic), and frequency (all three). For student characteristics, the effect of overall reading proficiency as measured by state test performance and the effect of being an English learner, were examined, in addition to controlling for grade.

Methods

Participants

Students were participants in the Word Generation project. The sample included 167 students and was balanced by gender, was ethnically diverse, included students with and without English proficiency and state test proficiency, and comprised students in sixth through eighth grades. These demographic characteristics are shown in Table 1. The sample of students was gender balanced, and the majority of students were classified as low socioeconomic status ($N = 136$; 81 %). Students were mostly classified as English speakers ($N = 120$; 72 %) with a smaller proportion classified as current or former English learners ($N = 47$; 28 %). Table 2

Table 1 Descriptive statistics for student abilities (N = 167)

	n	(%)	M	(SD)
Gender				
Female	81	(49)		
Male	86	(51)		
Race				
African-American	89	(53)		
Asian-American	14	(8)		
White	14	(8)		
Latino	49	(29)		
Multiethnic	1	(1)		
Student receives free/reduced price lunch				
No	31	(19)		
Yes	136	(81)		
Grade				
Sixth	66	(40)		
Seventh	60	(36)		
Eighth	41	(25)		
English proficiency level				
Limited	21	(13)		
Fluent	26	(16)		
English L1	120	(72)		
Proficient on MCAS				
No	67	(40)		
Yes	100	(60)		
Age			12.63	(1.02)
Number of clauses in essay			17.52	(6.52)
Total quality score for essay			11.23	(1.95)

L1 first language

provides descriptive statistics for the outcome variables, student ability predictors, and word characteristic predictors.

Word generation context

The Word Generation project as it was taught in this study was a weekly supplemental vocabulary program, taught for approximately 20 min per day by a variety of teachers. Each week's curriculum contained a high-interest topic that lent itself to a variety of potential opinions, and each week introduced and taught five vocabulary words from the Academic Word List (Coxhead, 1998). An entire middle school, grades 6–8, would complete the same curricular unit simultaneously, meaning all grades were focused on the same content each week, and reports from

Table 2 Descriptive statistics and bivariate correlations for variables in models

Variable	Scale	Mean	SD	Min	Max	Bivariate correlations[a]					
						1	2	3	4	5	6
Student × word outcome measures (N = 4175)											
1 An attempt to use a specific word (ATT)	Dich.	0.41	(0.49)	0	1	–					
2 Uses of a specific word for a student (USES)	Count	0.44	(0.55)	0	4	.95	–				
Student variables (N = 167)											
3 Grade 6 (GR6)	Dich.	0.85	(0.79)	0	1	.09	.09	–			
4 Grade 8 (GR8)	Dich.	0.25	(0.43)	0	1	–.02	–.04	–.46	–		
5 Proficient on MCAS (MCAS)	Dich.	0.6	(0.49)	0	1	.17	.16	.09	.01	–	
6 Fluent English proficient (FEP)	Dich.	0.13	(0.33)	0	1	.03	.03	.06	–.13	.16	–
7 Limited English proficient (LEP)	Dich.	0.16	(0.36)	0	1	–.08	–.08	–.01	–.13	–.19	–.16
Word variables (N = 25)											
3 Standard Frequency Index (FREQ)	Scale[b]	50.24	(4.68)	42.20	59.40	.09	.10	–			
4 Summed bigram frequency (BGF)	Count[b]	34,446	(15,942)	8212	57,568	.01	.00	–.36	–		
5 Morphological family size (MORFAM)	Count[c]	25.68	(20.74)	1	78	.03	.03	.50	–.47	–	
6 Imageability rating (IMG)	Scale[b]	3.22	(1.17)	1.94	5.65	.00	–.01	.47	–.42	.65	–
7 Phonological Levenshtein distance (PLD)	Mean edit distance[b]	3.18	(0.79)	1.75	4.35	–.04	–.04	–.58	.67	–.36	–.29

Dich. dichotomous

[a] Bivariate correlations are given for the transformed version of MORFAM. For correlations with ATTEMPTS and USES, the item level correlations (rather than aggregated versions of the variables) were used

[b] These variables were standardized (z-scored) for analysis

[c] MORFAM was transformed by taking the square root of the original metric and then standardized

teachers in the program indicate that the taught words were quite new to the majority of students. The words selected were general academic terms that could be used in a variety of contexts, not specific disciplinary terms heavily related to the topics. Students had a variety of opportunities throughout the week to see the words, hear them spoken, discuss their meanings, and use them to discuss opinions on the weekly topic, and the culminating activity of the week was a short opinion essay handwritten by students stating their opinions (see Appendix 1 for a full list of essay prompts and taught words). The writing was short in length (descriptive information about essay length in clauses is included in Table 2 for reference), as it was generated in approximately 20 min, and responses were unedited rough drafts. Writers were verbally encouraged to use the words in their essays by their teachers.

No data from the school were available about student prior knowledge or exposure to the particular words taught as a part of this curriculum prior to the start of the program. Though an overall achievement measure was included, there was no way to be certain how much receptive knowledge of the 25 taught words students had prior to the program, and we are assuming that students have varying exposure. Though the overall achievement measure does contain a measure of expressive writing skill, we are unable to isolate this particular subcomponent, leaving a general representation of overall literacy skill, but no specific measure of expressive written language. Because the phenomenon of interest here is in students choosing to use words in writing and not focused on growth in receptive word knowledge, the focus is on these expressive uses in this study.

Prior studies using this dataset used an informal measure of writing quality, a rubric used by the school and the state achievement test to measure overall quality (basic descriptive information included in Table 2 for reference). This study showed a significant relationship between overall writing quality and attempts (including incorrect or partially correct uses) at using general academic language markers in the essays (Dobbs, 2014). However because using academic language is a key element of writing quality, overall writing quality is not used as a measure in this study, because of its confounded relationship to academic vocabulary use.

Measures

Outcome: attempts (ATTEMPTS) and uses (USES)

The outcome variable was whether students made any attempt to use a target word, regardless of whether the word was used correctly. Thus, there were 4175 total cases in the data set, one for each student ($N = 167$) and each word ($N = 25$). Word uses were coded in two ways, which allowed us to examine two slightly different outcomes. Words were first coded by *attempts*. For this coding system, a student received a value of 0 if they never used a word and a value of 1 if they ever used a word, regardless of the number of times they used the word. Overall, 1715 uses by students were recorded, with considerable variability across words (range = 40 [*inevitable*] to 104 [*function*]) and students (range = 0–24; $M = 10.99$, $SD = 6.76$).

Words were also coded by *total uses*. For this system, a student received a value equal to the number of times a particular word was used across all essays the student wrote. Students tended not to use the same word very many times. There were 1623 cases of a single use, 81 cases of two uses, 10 cases of three uses, and just 1 of four uses (for one student on *project*). For the coding of both *attempts* and *uses*, a misspelled word was counted as an attempt to use a particular target word (Table 3).

Importantly, we concern ourselves with any use at all in this study for several reasons. First, the goal was a coding scheme that captured the complexity of students attempting to use newly learned items in writing, which often reflect partial and developing knowledge of words. The phenomena of interest was factors that might influence word use, including in these cases of partial understanding. Second, students in this study had been working with words for only a single week at the time of these writing samples, a very short span in which to develop entirely correct productive uses of new vocabulary items. It should be noted that Perfetti (1992) has

Table 3 Results for ATTEMPTS and USES models

Variable	Main effects ATTEMPTS (binomial)		Main effects USES (continuous)	
	Coeff.	(SE)	Coeff.	(SE)
λ_0 Intercept	−1.227	0.264***	0.061	0.048***
Student				
γ_{010} GR6 (Grade 6)	0.457	0.258[†]	0.094	0.046*
γ_{020} GR8 (Grade 8)	−0.085	0.317	−0.008	0.053
γ_{030} MCAS (state test proficiency)	1.064	0.235***	0.164	0.041***
γ_{040} LEP (limited english proficiency)	−0.687	0.326*	−0.104	0.056*
γ_{050} FEP (fluent english proficiency)	−0.208	0.346	−0.040	0.061
Word				
γ_{001} FREQ (standard frequency index)	0.390	0.155*	0.067	0.155*
γ_{002} BGF (bigram frequency)	0.002	0.162	0.025	0.162
γ_{003} MORFAM (morphological family size)	0.930	0.151	0.043	0.151
γ_{004} IMG (imageability)	−0.182	0.137	−0.053	0.028[†]
γ_{005} PLD (phonological Levenshtein distance)	−0.151	0.178	−0.001	0.179

Random effect	Variance			Variance
r_{010} Student	1.891			0.061
r_{001} FREQ	0.276			0.010
r_{002} PLD				0.000
r_{001} Word	0.208			0.010
r_{002} GR8	0.246			
Residual				0.206

[†] $p < .10$; * $p < .05$; ** $p < .01$, *** $p < .001$

described orthographic output as a strong test of the quality of individuals' bonded orthographic, phonological, and semantic representations, so we believe this measure provides considerable insight about students' nascent knowledge of these words.

However, neither measure ultimately concerns whether the uses were correct. An analysis on correct uses presented a challenge to our question here, about factors likely to influence the attempt to use new items. This analysis does not include correct uses because a score of 0 could be interpreted as either an incorrect use or the lack of a use. These factors could not be separated and thus we opted to consider attempts and uses regardless of accuracy, as it is more central to our question of interest. Given this limitation, a qualitative exploration of these factors in the results is provided.

Student abilities

We collected several types of child data, namely their proficiency on the state English Language Arts test, their English learner status, and their grade. For these variables, there was very little missing data. There were, however, four participants who had used some words (a total of 12 uses) but for whom we were missing data on the other participant characteristics. In these cases, used list-wise deletion, that is, we deleted all data for a participant if they had missing scores on MCAS or did not have EL status coded.

Proficiency on state testing (MCAS)

Students' proficiency on the Massachusetts Comprehensive Assessment System (MCAS) test of yearly progress was an important student ability. It was expected that students who were proficient on the MCAS would be more likely to use the vocabulary words being taught, as they would likely have better facility at learning new words than their peers with lower scores (Perfetti, 2007). MCAS proficiency was a dichotomous variable.

English learner (EL) status (FEP and LEP)

Students' status as an English learner (EL) was included in the analysis. It was expected that students learning English would be less likely to use vocabulary words in their writing. Schools provided data on whether students were currently classified as limited English proficiency (LEP) or formerly classified as limited English proficiency (FEP). The school district used the Massachusetts English Proficiency Assessment (MEPA) and additional information to classify students, but their MEPA scores at the time of this study were unavailable. Therefore school classification status was used to determine whether they were language learners. Both the LEP and FEP status variables were dichotomous, where a student whose first language is English would receive 0 for both LEP and FEP.

Grade (GR6 and GR8)

Students' grade was also included as a predictor. Seventh grade was used as the reference category and included dichotomous variables for sixth grade (GR6) and eighth grade (GR8). We expected a positive effect for the GR6 variable because the sixth grade students tended to write more than the older students and thus had more opportunities to use the target words. An ordinal grade scale was not used because linear change due to grade itself was not expected.

Word characteristics

Bigram frequency (BGF)

Bigram frequency was derived from the data available from the English Lexicon Project (Balota et al., 2007). The bigram frequency measure used was the summed bigram frequency, which included frequencies for every bigram in the word, such that longer words have higher bigram frequency. This measure's correlation with mean bigram frequency was .79 and its correlation with length was .72. It was expected that words with higher bigram frequencies would be attempted more often because they contained more familiar orthographic patterns which could make spelling easier.

Frequency (FREQ)

Frequency was measured using the standard frequency index (SFI) reported in the *Educator's Word Frequency Guide* (Zeno, Ivens, Millard, & Duuvuri, 1995). SFI values are transformed from the U statistic. U represents a word's type frequency per million tokens, adjusted for the dispersion across content areas. Breland's (1996) formula for SFI is as follows: $SFI = 10 \times (\log_{10}U + 4)$. The word SFI ranged from 39.9 to 59.4, with a mean of 50.56 ($SD = 4.77$). Among words appearing in the sixth, seventh, or eighth grade corpus, the mean was 47.55 ($SD = 6.29$). Thus the dispersion of word frequencies was only slightly smaller than that for many words appearing in texts at these grades.

Imageability (IMG)

Imageability is an index of whether a word has a meaning that can be mapped to a mental image. It is thought to indicate the degree to which semantic processing affects word processing. Although there are not writing imageability data, Cortese and Khanna (2007) and Cortese and Schock (2013) observed that imageability affected adults naming and lexical decision accuracy response times and ratings, for monosyllabic and disyllabic words, respectively. Importantly Cortese and Schock (2013) observed imageability effects for polysyllabic words when accounting for age of acquisition, while Cortese and Khanna (2007) did not. This aligns with the idea that these words may require more semantic activation that monosyllabic ones, similar to Kearns (2015) finding of a vocabulary size effect for word naming

accuracy in elementary-age students. In the present study, the hypothesis was that students would be more likely to attempt to write more imageable words, both because they understood their meanings and because they would be able to use them in writing sentences in semantically and syntactically logical ways than less imageable words. The imageability data for the present study were collected as part of a larger experiment with undergraduate and graduate students at Boston University. Twenty students were given instructions and rated the randomly ordered words—interspersed with unrelated words—following the procedures of Cortese and Fugett (2004) but substituting the mascot of the students' institution for the University of Colorado mascot. The internal consistency of the ratings was .92, strong particularly given the size of the sample. We had confidence that the ratings reflected adults' general perception of words' imageability.

Morphological family size (MORFAM)

The word's morphological family size was the number of words that had the same root word as the target word. For example, the morphological family size for *sufficient* is 6, which represents the number of English words that are morphologically related to *sufficient*, namely *insufficiency*, *insufficient*, *insufficiently*, *sufficiency*, *sufficient*, and *sufficiently*. Morphological family size has been shown to relate to word recognition in developing readers (Kearns et al., in press; Carlisle & Katz, 2006; Carlisle & Stone, 2005). A type-based measure of family size is thought to be superior to a token-based measure (Reichle & Perfetti, 2003; Schreuder & Baayen, 1995). In the present study, we expected that morphological family size might influence readers' likelihood of attempting to use a word. This might even create cases of readers attempting to use words but doing this incorrectly. To establish the morphological family size, every derived or inflected word from the same root was counted using the morphological coding in the English Lexicon Project (ELP; Balota et al., 2007) unrestricted database. Compound words were included in a word's morphological family size count. The unrestricted version of the ELP database contains 79,672 English words and information about roots and affixes for 68,624 of these.

Phonological Levenshtein distance (PLD)

A word's phonological Levenshtein distance (PLD) is a measure of how phonologically unique the word is. Levenshtein distance is a form of edit distance, which counts the number of changes (additions, substitutions, etc.) that must be made to a word to locate its 20 nearest neighbors, that is, the 20 words that are most phonologically similar to it. For example, *cycle* has a PLD of 1.75, meaning that the 20 nearest neighbors might require a change of only one or two letters (e.g., *cycled* and *cycles* require only one phonological change each, the addition of /d/ and /z/ respectively). For polysyllabic words, PLD has been shown to predict more variability in adults word naming and lexical decision response times than the more commonly used phonological *N,* an index of the number of words that can be created with just one phonological change (Yap & Balota, 2009, but see Cortese &

Schock, 2013). For polysyllabic words, phonological N is often zero, but variability in PLD will still be present. In the present study, the correlation between PLD and phonological N is $-.63$, but phonological N is 0 for 18 of the 25 words.

Analyses

For the analyses, the scales of the variables affect interpretation of the intercept. Because dichotomous covariates for sixth and eighth grade were included, the intercept reflected performance for a seventh grader. In addition, the dichotomous scaling of LEP, FEP, and MCAS meant that the intercept was for a student whose first language is English and who was not proficient on MCAS. The estimation procedure, described below, has difficulty handling variables with very different scales, so the continuous outcome (USES) and the continuous predictors were all transformed to z-scores. To assure that standardization was appropriate, the normality of each variable was checked, and the morphological family size (MORFAM) predictor was transformed by taking its square root, a transformation selected because it improved normality versus the untransformed variable. The correlation matrix for the ATTEMPTS and USES outcomes, grade, dichotomous predictors, standardized predictors, and the transformed and standardized MOR-FAM predictor is given in Table 2.

In the language of multilevel modeling (e.g., Raudenbush & Bryk, 2002), this is considered a cross-classified random-effects model. For the binary model, we use the term "explanatory item response model," consistent with the language used by De Boeck and other methodologists who developed the *lme* function used for these models. The main analytical tool for the binary analysis was an explanatory item response model (De Boeck & Wilson, 2004), specifically, an item response model with a random item parameter (De Boeck, 2008; Janssen, Schepers, & Peres, 2004; van den Noortgate, de Boeck, & Meulders, 2003).

The explanatory item response model was constructed to reduce the variability in person abilities and item difficulties with person characteristics (student writing-related abilities), and item characteristics (word characteristics). For the attempts (ATTEMPTS) analysis, Laplace approximation implemented in the lmer function (Bates et al., 2015) from the lme4 library in R was used. Researchers have tested the effects of student abilities and item characteristics on first graders' ability to pronounce nonwords (Gilbert et al., 2011); third and fourth graders' ability to pronounce polysyllabic words (Kearns, 2015); fifth graders' ability to pronounce polymorphemic words (Kearns et al., in press) and middle schoolers' ability to pronounce, spell, and indicate familiarity with polysyllabic polymorphemic words (Kearns & Al Ghanem, 2014; Goodwin, Gilbert, & Cho, 2013); and Kindergarten-ers' ability to say letter sounds (Piasta & Wagner, 2010). To our knowledge, this study is the first to evaluate characteristics of students' writing using these models.

For the word uses (USES) analysis, data were analyzed using cross-classified random effects models for continuous outcomes using lmer with maximum likelihood estimation. The estimation procedure is different than that for binary outcomes, but the variability is partitioned in the same way. Unlike with binary

outcomes, there is also a residual term representing the variance unrelated to student or word.

For both types of analyses, three models were created to answer the research questions. First, an *unconditional model* containing only an intercept and random effects was created. Second, a *main effects model* was used to evaluate the person and item characteristics that reduced person and item variability. To evaluate the reduction in variability, two methods were used. First, we calculated intraclass correlations reflecting the reduction in variance due to student and item covariates. In addition, 95 % plausible values ranges were used; the reduction in these ranges provides a qualitative sense of variance reduction. The methods used to build these models and calculate the plausible values range have been detailed elsewhere (e.g., Kearns, 2015; Baayen, Davidson, & Bates, 2008; Gilbert et al., 2011), and the description is not replicated here. The method used to determine the correct random effects structure followed the recommendations of Bates (2011) and most closely resembles that followed by Kearns (2015). The approach differed from that of Kearns only in that an approach used for binary data were applied to a continuous model. The main effects model is given in Appendix 2.

Results

Attempts (ATTEMPTS) analyses

For the attempts (ATTEMPTS) analyses, in the unconditional model including random effects for student and word, the mean probability of an attempt to use a word was .37 for an average word and an average student in the study. The 95 % plausible values range for students was .03 to .91, indicating that, for an average word, there was a 95 % chance a student's likelihood of a use would fall within that range. The 95 % plausible values range for words was .15 to .65, indicating that, for an average student, there was a 95 % chance that a student's likelihood of attempting to use a word would fall within that range. The intraclass correlation (ICC) for student, conditional on word, was .38, and the ICC for word, conditional on student, was .09. These plausible value ranges and ICCs indicate that much of the variability in students' likelihood of choosing a correct word was within the student, rather than within the word. However, the amount of variability due to word was enough to consider explanatory variables. In addition, a simpler model without a word random effect fit worse than the model with word, $\Delta\chi^2_{[1]} = 173.27, p < .0001$, so word variability could be modeled.

For the main effects model, addition of the 5 word fixed effects and 5 student fixed effects improved the fit of the model over the unconditional model, $\Delta\chi^2_{[10]} = 38.43, p < .0001$. In addition, tests of random word slopes showed better fit with a GR8 slope that correlated with the intercept. This indicated that eighth grade students had a different pattern of attempts to use words than seventh grade students. Tests of random student slopes showed better fit with a FREQ slope that correlated with the intercept. This indicated that the effect of frequency differed across students, with some students being more affected by the frequency of the

items and others less. No other random slopes improved model fit when entered alone, and the model with the additional slopes and correlations fit better than the main effects model without them, $\Delta\chi^2_{[4]} = 67.12$, $p < .0001$. The mean probability that a student would use one of the words was .22. All probabilities hereafter are given for an otherwise-average seventh grader with an English L1 and without MCAS proficiency, except as noted. The 95 % plausible values range for students was .03 to .87 and the 95 % plausible values range for words was .15 to .53. Comparing the plausible values ranges for this model and the conditional model indicates that the main effects model reduced the variability but that considerable additional variability remained. Taking the intercept values from the unconditional model and the main model, the model reduced the variability in students by 5.7 % and the variability in words by 57 %.

For the variables of interest in the main effects model, three predictors were related to the likelihood that a student would use a particular word. For student abilities, there an MCAS effect, $\hat{\gamma}_{030} = 1.06$, $\Delta\chi^2_{[1]} = 19.65$, $p < .0001$. This effect meant that an otherwise-average student who was proficient on the MCAS had a .46 probability of attempting to use an average word, versus .22 for an otherwise identical student. There was also a LEP effect, $\hat{\gamma}_{040} = -0.687$, $\Delta\chi^2_{[1]} = 4.38$, $p = .04$, such that a student with limited English proficient had a .13 probability of attempting to use a word compared with an otherwise identical student whose L1 was English, whose probability was .22. There was a marginally-significant effect for GR6, $\hat{\gamma}_{010} = -0.687$, $\Delta\chi^2_{[1]} = 3.09$, $p = .08$ that suggested that sixth graders attempted more words (probability of .32 versus .22), but it was not reliable. There was one significant word effect, that of FREQ, $\hat{\gamma}_{001} = 0.390$, $\Delta\chi^2_{[1]} = 5.78$, $p = .02$, such that an average student, on a word with an SFI value 1 *SD* above the mean (SFI = 55.3; e.g., *indicate*) had a probability of attempting the word of .30 while the same student, on a word with an SFI vowel 1 *SD* below the mean (SFI = 45.8; e.g., *initiative*), had a probability of making an attempt of .17.

Word uses (USES) analyses

The USES analyses produced exactly the same pattern of effects as the ATTEMPTS analyses, although the outcome was slightly different, reflecting the predicted number of uses rather than the probability of correct use. For the unconditional model, the student ICC was .13 and the word ICC was just .03, suggesting that much of the variance in the number of uses was noise not associated with student or word factors. Nonetheless, models without the word and student random effects fit the data worse, so these word and student random effects were retained and the main effects models including word and student covariates examined. These small ICCs meant that differences in the number of times a student attempted to use a word are related to the student's experience with a specific word. These are factors probably related to the immediate instructional context that could not be observed here (e.g., whether the student was paying attention during the explanation of a given word).

For the USES main effects model, there were PLD and FREQ slopes for the random student effect and a correlation between the FREQ slope and random student intercept. The main effects model reduced the student variance by 14.1 %

and the word variance by 35.5 %. The USES fixed effects in the main effects model followed the same pattern as the ATTEMPTS ones. There was a significant MCAS effect, $\hat{\gamma}_{020} = 0.164$, $\Delta\chi^2_{[1]} = 15.01$, $p < .0001$. This meant that, for a given word of average difficulty, an average student with MCAS proficiency would use the word 0.17 more times than a student without it. Across the entire set of 25 words, that represents more than 4 more word uses. The LEP effect, $\hat{\gamma}_{030} = -0.104$, $\Delta\chi^2_{[1]} = 3.35$, $p = .07$, was only marginally significant. A LEP student was likely to use a word 0.11 fewer times than an L1 English peer, representing nearly 3 fewer word uses across the 25 words, although this was not a reliable effect across both analyses. Finally, there was a FREQ effect, $\hat{\gamma}_{001} = 0.067$, $\Delta\chi^2_{[1]} = 4.55$, $p = .03$. A word 1 *SD* below the mean received 0.06 fewer word uses than an average word, and a word 1 *SD* above the mean received 0.06 more uses. This means that infrequent (-1 *SD*) words might be used 3 fewer times than frequent words ($+1$ *SD*), across the 25.

Sample uses from the data

The quality of individual word uses in this study were not examined, given the reasons already explained. Our findings concerning attempts, therefore, provide no direct evidence of the relation between attempts to use the words and the quality of word uses. However, we are interested in the quality of students' attempts to use words, so some qualitative examples from students' samples to try and understand patterns in whether readers were using high frequency words correctly are examined. These samples were identified by coders with the aid of CLAN software, and then coded by a team of two coders who were trained by the first author and achieved high levels of reliability. These will be discussed below.

Frequency and quality

One student example of a strong use for a higher frequency word, *function*, shows an effective use of the word: "The primary *function* of school is to make sure that all students have specific knowledge about history, science, literature, and mathematics." The sample contains a sound assertion about the purpose of school. A simpler but still appropriate use of *function* was this: "The primary *function* of school is learning new subjects." This use is similar to the first, although the phrase "learning new subjects" is less clear than "have specific knowledge" in the first example. A third case shows a less clear sentence—but not incorrect attempt to use *function*: "The main *function* of school is to get job so you can live becuse[1] nothing is free." This use indicates an understanding of a distal function of school but misses the primary direct function, getting an education (that would ultimately make it easier to get a job). An important point, however, is that these uses share understanding of some lexical features of a higher frequency word, albeit with varying levels of incorporation into a broader argument.

---

[1] All sample text from students is transcribed as written by students without editing for spelling, capitalization, or punctuation.

One rarely-used word, *manipulate,* was used in the following way by a student in response to a prompt about whose responsibility it is when adolescents take up smoking: "Tobacco companies target teens and *manipulate* them with commercials and things that attract them." There is evidence here of a high-quality lexical representation in that it is collocating properly and reflects an understanding of multiple semantic features (e.g., coercive; clandestine; using the target's preferences). The orthographic representation is also high in quality, and it is spelled correctly.

Thus, frequency appears to affect the quality of a word use and may index the overall quality of the lexical representation, but individual learner knowledge of the words involves factors beyond the likelihood they had encountered the word that frequency describes.

MCAS proficiency and quality

As our results show, students with MCAS proficiency were more likely to attempt words in general, and their uses were more likely to be high in quality. A student with a proficient score used the word *inevitable* in the following way, when arguing who is responsible if teens take up smoking: "Many doctors belive it *inevitable* that smoking will eventually lead to an accumulation of health concerns." This reflects a high quality representation of the word in an unusual syntactic construction with proper collocation and a clear relationship to the rest of the material in the sentence.

By contrast, a student without MCAS proficiency used an unfamiliar word, *inevitable* in this way: "Another reason is if teens or anyone keeps somking they'll *ineviiable* canse [cancer] like it says in paragrah 1." Here is a much less robust lexical representation of *inevitable* than the use of *manipulate* just described. The use shows the association between smoking and a long-term very likely result, cancer. But, the syntax of the sentence is incorrect and several words that would clarify understanding are missing. Moreover, the word was used in a very similar way in the anchor text, suggesting that this use may reflect a very shallow—and possibly domain-specific—understanding of the word. Finally, the word is also spelled incorrectly, suggesting a lower quality orthographic representation for this word (and many others).

These examples provide some evidence that MCAS proficiency relates to quality in addition to the number of attempts to use the word. This supports the idea that lexical quality involves both word factors (frequency) and learner factors (general reading and writing proficiency).

Discussion

We begin with a brief summary of the results. First, counting attempts and counting the number of uses of each word attempted produced quite similar patterns of effects. Second, for the attempts analysis, it was observed that frequency affected the probability of an attempt, with higher frequency words being more likely to be used. Students with a LEP designation were less likely to use words than their L1

English peers, and students proficient on MCAS were more likely to use words than their peers who were not proficient. The results suggested that students used frequent words more often than infrequent ones and that students proficient on MCAS tended to use more words than their less proficient peers.

Frequency effects in word use

The frequency effect—and the absence of other effects—may have a simple interpretation within a lexical quality framework. In short, frequency may index how likely it is that a reader has orthographic, phonological, and semantic information about the word, even before it is taught. The domain-specific lexical quality measures may have been superseded by a frequency effect that reflected the possibility of high lexical quality across all three domains.

Why would frequency have this effect? One possibility is that word frequency partly indexes the *utility* of a given word. Higher frequency words tend to be used in a wide variety of texts (Hiebert, 2012; Zeno et al., 1995), so writers may have found it easier to use these words in their writing. A case in point is *function*, the word most frequently attempted. The students might have found it especially easy to include *function* in their writing because it would be easily incorporated in their texts without linking the word to a given context. By contrast, the rarely-used word *inevitable* has limited utility, being only useful in very specific contexts.

Another possible interpretation is that students could have built lexical representations for the words both before the intervention and within it. Before the intervention, the students would likely have experienced some of the higher frequency words, even if they only acquired limited representations of them (e.g., some memory for the letters, a partial pronunciation, and a few semantic features). Within the intervention, the instructional materials incorporated newly taught vocabulary words, and higher frequency words were used more often in the materials. The teachers were also more likely to use the higher frequency words more often, giving students more opportunities to build the quality of the representation. Thus, the frequency of the words affects the likelihood of acquiring a high-quality orthographic, phonological, and semantic representation of the word because higher frequency words would likely have been used before and during the intervention. This interpretation suggests that the lexical quality hypothesis—which concerns reading—could extend into productive vocabulary in writing, with students forming more robust representations of higher frequency words leading to more attempts at use.

Learner effects in word use

We also unsurprisingly find that learner characteristics related to the likelihood that students would attempt to incorporate a newly taught word into writing. Students' Language Arts MCAS results were used as a proxy for general literacy skill, and students with higher proficiency on the MCAS were found to be more likely to write new vocabulary words into their essays. The MCAS measures both reading and writing skills, so this finding likely bears on both competencies. In terms of reading,

these findings align with results concerning adolescents' ability to understand words while reading: Better readers understand more individual words; better readers were also found to be more likely to use newly taught vocabulary words. This may reflect the overall greater vocabulary size of these readers, meaning that they are able to incorporate more words because they are more likely to have representations for them.

With regard to writing, good writers have strong language skills more generally, which is particularly helpful given the cognitive complexity of writing tasks. For stronger writers, many facets of writing skill have become more automatic (e.g., spelling) (Kellogg, 2008), leaving more cognitive space for composing thoughts or attempting to incorporate words with less robust lexical representations into text. Therefore, as we see in this study, students who have stronger literacy skills are at a sort of dual advantage, with stronger ability to build lexical representations for new items and stronger writing skills to allow for the cognitive demand of incorporating new vocabulary items into text.

There is also a possibility that MCAS performance links to other learner abilities not captured in our data. For example, better readers and writers may have greater task persistence, inhibitory control, or ability to follow directions, all of which may have increased the likelihood the students incorporated these words. It is most likely that all three factors are at work here, similar to what Ricketts et al., (2007) argued in the case of the relation between vocabulary and word reading. Put simply, better readers and writers are well positioned to try to use new vocabulary in unstructured composition, a pattern which was borne out as qualitative sample uses of words were pulled from our data.

Future research and limitations

These findings raise interesting questions about future research and instructional approaches. First, there is much to be learned about the ways that semantic, orthographic, and phonological information influence whether learners use newly-learned vocabulary in their writing. One way to study this might be to have students write spontaneous compositions—as they did for us—but to require (rather than encourage) uses of a fraction of the words. Researchers could then examine more carefully what factors might lead students to choose some words over others. To examine quality, students could be asked to write sentences or short paragraphs containing taught vocabulary words and to evaluate how well the vocabulary words were used. Then, as in this study, examination could be conducted about the word and learner characteristics that contribute to word use, including a variety of word characteristics to further determine whether frequency effect remains similar or whether other characteristics also become significant. An extension of this work might be to include measures of students' item-specific knowledge as predictors of the quality of vocabulary uses. For example, with word reading, Kearns (2015) examined whether students' ability to pronounce letter-sounds, phonograms, affixes, and roots in specific words (e.g., for *scientist*, its letter-sounds, phonograms like *en*, the affix *–ist,* and the root *science*) affected their ability to read the polymorphemic word (e.g., *scientist*) correctly. Similar procedures—albeit measuring constructs

more closely linked to dimensions of lexical quality (cf. Kearns & Al Ghanem et al., 2014)—could be used to examine the quality of written vocabulary word uses.

Examining the intersection between the instructional environment and word use (quality) represents another important focus for future research. By coding factors like the number of times teachers used the words, how clearly the teachers defined and explained the words (e.g., using tools such as Frayer models), how often they occurred in text, and the richness of the text contexts would also be valuable to explore.

Several limitations should be considered. First, the amount of information varied considerably across grades. Fewer eighth graders participated in the program regularly, compared with sixth graders. This is one reason that there appears to be a negative effect for eighth grade on the likelihood of use. We did, however, take advantage of all available data. In other words, we were able to use any cases where an eighth grader did participate.

Second, analysis is limited here to students' attempts to use newly learned academic vocabulary items, and we do not delve here into correct uses of these items because of the duration of the intervention, the intent to capture partial knowledge of new items, and the analytic challenges of parsing incorrect uses from essays lacking attempts. Because of the short duration of this intervention, it was not expected that students would have acquired fully robust representations of newly taught items after 1 week of short sessions working with the words. Now that some word and learner factors of interest have been generated, it would be useful to ask students to write using new items following a longer time span to develop word knowledge. It would also be useful to develop tools that could parse the continuum of expressive word knowledge to better capture partial understandings of words, and in the writing of these novice academic language users, capturing incorrect, partially correct, and correct word uses is certainly a further question of interest.

There are also differences due to student, word, and teacher factors that we cannot explain. Limited data were available about student skills beyond the MCAS. As described in our recommendations for future research, it also would be helpful to know how often teachers were able to incorporate the words into speech over the course of a week. It stands to reason that less frequent words might be more difficult to incorporate into everyday speech, but it is unclear whether this pattern was borne out in the classrooms here. There is also likely an interaction between teacher quality or investment in the Word Generation program and the likelihood a student would use a word. This is one reason that much of the variability in the uses analysis was related neither to students or items specifically—only 13 % was related to students and 3 % to words. This means that most variability was related to contextual factors that would affect the number of times a student might use a word. For example, students may have learned more about some words than others through instruction based on instructional quality, such that their uses of a particular word depended not only on the word's frequency or the student's academic skills but on how well the student acquired an understanding of the specific word sufficient to use it one or more times in text. Another possibility is that the particular writing task assigned for a given word might affect how often they used that word compared to the task given for another word.

In addition, we could not control for student exposure prior to participation in Word Generation, as the broader study was not initially set up to collect a measure of prior knowledge of these particular words. Prior exposure is doubtlessly important; reading researchers have repeatedly underscored its critical role in the acquisition of orthographic representations (e.g., Share, 1995). Moreover, as Perfetti (1992) pointed out, orthographic output is an important indicator of a strong bonded representation. So, the absence of controls for prior knowledge means that we cannot be sure whether the frequency effect is general or due to readers' word-specific representations. We assume that it is item-specific, in alignment with item-based accounts of reading acquisition, but controls for prior exposure would be helpful to make this clear. Future research could extend this line of research by collecting data on students' prior knowledge of the particular words being taught in a vocabulary intervention.

Additionally, connections to writing quality were not made in the analysis here, because the relationship between word and learner characters and the likelihood of use was the central question. The rubric used for overall writing quality in prior research with this dataset proved too general to help us illuminate academic vocabulary use. But it would be worthwhile to continue developing quality tools to help us understand this relationship, and this line of research could be extended by finding ways to extend these methods to model the relationship between writing quality and the use of particular newly learned vocabulary items.

These issues are one reason that analysis did not explain more of the variability in students and words. We simply did not have the data to examine environmental variables that were certainly related to word use. Given this limitation, it is interesting that even a frequency effect was observed. While the lack of contextual information places limits on our ability to generalize, we wish to emphasize that these factors largely add noise to the data and circumscribe the likelihood of finding effects. In other words, these effects are possibly meaningful because they emerged despite our lack of ability to model conditions that would have made it easier to detect effects.

Conclusion

Despite these limitations, this study is unique in several ways that make it interesting to readers and valuable to the field. First, this paper is one of the first to use cross-classified random effects models to examine word use in writing. This approach has other possible applications in writing research, and we hope writing researchers will see the potential value of the approach, given these findings. Second, factors have been identified that may provide a path toward better understanding what explains whether students try to use new vocabulary (and eventually, how well they do this). These results provide tentative support for the idea that lexical quality is important in word choice in writing. Finally, descriptive studies like this will help the field identify ways of improving vocabulary instruction that can be tested with causal models.

Appendix 1

Topics/essay prompts selected for the data corpus

Week #	Essay prompt	Words taught from the academic word list
1	What is the purpose of school?	*Analyze*
		Factor
		Function
		Interpret
		Structure
5	Should hip hop music be censored?	*Considerable*
		Contribute
		Demonstrate
		Sufficient
		Valid
8	Who is responsible for stopping global warming?	*Attribute*
		Cycle
		Hypothesis
		Project
		Statistics
12	Should junk food be sold in schools?	*Acknowledge*
		Incidence
		Incorporate
		Initiative
		Transport
15	Who is responsible for the problem of teen smoking?	*Accumulation*
		Contradict
		Exhibit
		Inevitable
		Manipulate

Appendix 2: ATTEMPTS model

Unconditional model

Level 1 (Response$_{jkmi}$)

$$\text{logit}\left(p_{jkmi}\right) = \lambda_{0ji}$$

Level 2 (Person$_{jkm}$ & Item$_i$)

$$\lambda_{0ji} = \gamma_{00} + r_{001j} + r_{002i}$$

$$r_{001i} \sim N\left(0, \sigma_j^2\right)$$

$$r_{002i} \sim N\left(0, \sigma_i^2\right)$$

Student and Item model

Level 1 (Response$_{jkmi}$)

$$\text{logit}\left(p_{jkmi}\right) = \lambda_{0ji}$$

Level 2 (Person$_j$ & Item$_i$)

$$\lambda_{0ji} = \gamma_{000} + \gamma_{010}GR6_j + (\gamma_{020} + r_{002})GR8_j + \gamma_{030}MCAS_j + \gamma_{040}LEP_j$$
$$+ \gamma_{050}FEP_j + (\gamma_{001} + r_{002})FREQ_i + \gamma_{002}BGF_i + \gamma_{003}MORFAM_i$$
$$+ \gamma_{004}IMG_i + \gamma_{005}PLD_i + r_{010j} + r_{001i}$$

$$\begin{bmatrix} r_{010} \\ r_{020} \end{bmatrix} \sim MN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{r010}^2 & \sigma_{r010,r020}^2 \\ \sigma_{r010,r020}^2 & \sigma_{r020}^2 \end{bmatrix}\right)$$

$$\begin{bmatrix} r_{001} \\ r_{002} \end{bmatrix} \sim MN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{r001}^2 & \sigma_{r001,r002}^2 \\ \sigma_{r001,r002}^2 & \sigma_{r002}^2 \end{bmatrix}\right)$$

References

Andrews, S., & Bond, R. (2009). Lexical expertise and reading skill: Bottom-up and top-down processing of lexical ambiguity. *Reading and Writing, 22*, 687–711.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459.

Bates, D. (2011, June 22). *[r-sig-ME] Correlated random effects in lmer and false convergence*. Message posted to https://stat.ethz.ch/pipermail/r-sig-mixed-models/2010q2/003921.html.

Bates, D. M., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2015). Linear mixed-effects models using Eigen and S4 (Version 1.1-7).

Baumann, J. F., Kame'enui, E. J., & Ash, G. E. (2003). Research on vocabulary instruction: Voltaire redux. In J. Flood, D. Lapp, R. Squire, & J. M. Jensen (Eds.), *Handbook of research on teaching the English language arts* (2nd ed., pp. 752–785). Mahwah, NJ: Lawrence Erlbaum.

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

Biemiller, A. (1999). *Language and reading success.* Cambridge, MA: Brookline.

Bonin, P., Fayol, M., & Gombert, J. E. (1997). An experimental study of lexical access in the writing and naming of isolated words. *International Journal of Psychology, 33*, 269–286.

Breland, H. M. (1996). Word frequency and word difficulty: A comparison of counts in four corpora. *Psychological Science, 7*(2), 96–99.

Carlisle, J. F., & Katz, L. A. (2006). Effects of word and morpheme familiarity on reading of derived words. *Reading and Writing, 19*, 669–693.

Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly, 40*, 428–449.

Corson, D. (2002). The learning and use of academic English words. *Language Learning, 47*, 671–718.

Cortese, M. J., & Fugett, A. (2004). Imageability ratinge for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36*, 384–387.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *The Quarterly Journal of Experimental Psychology, 60*, 1072–1082.

Cortese, M. J., & Shock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *The Quarterly Journal of Experimental Psychology, 66*, 946–972.

Coxhead, A. (1998). *An academic word list.* Wellington, NZ: Victoria University of Wellington.

De Boeck, P. (2008). Random item IRT models. *Psychometrika, 73*, 533–559.

DeBoeck, P. & Wilson, M. (2004). *A framework for item response models.* New York: Springer.

Dobbs, C. L. (2014). Signaling organization and stance: Measuring the use of academic language markers in middle grade persuasive writing. *Reading and Writing, 27*(8), 1327–1352.

Duin, A. H., & Graves, M. F. (1987). Intensive vocabulary instruction as a prewriting technique. *Reading Research Quarterly, 22*, 311–330.

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General, 120*, 190–202.

Fayol, M. (2012). Cognitive process of children and adults in translating thought in written language in real time: Perspectives from 30 years of programmatic cognitive psychology and linguistics research. In V. W. Berninger (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. 27–59). New York: Psychology Press.

Flinspach, S., Scott, J., & Vevea, J. (2010). Rare words in students' writing as a measure of vocabulary. In R. Jiminez, V. Risko, M. Hundley, & D. Rowe (Eds.), *59th Annual Yearbook of the National Reading Conference* (pp. 187–200). Oak Creek, WI: National Reading Conference.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*, 365–387.

Ford, M. A., Davis, M. H., & Marslen-Wilson, W. D. (2010). Derivational morphology and base morpheme frequency. *Journal of Memory and Language, 63*, 117–130.

Gersten, R., Dimino, J., Jayanthi, M., Kim, J. S., & Santoro, L. E. (2010). Teacher study group impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal, 47*, 694–739.

Gilbert, J. K., Compton, D. L., & Kearns, D. M. (2011). Word and person effects on decoding accuracy: A new look at an old question. *Journal of Educational Psychology, 103*, 489–507.

Goodwin, A. P., Gilbert, J. K., & Cho, S. J. (2013). Morphological contributions to adolescent word reading: An item response approach. *Reading Research Quarterly, 48*, 39–60.

Hersch, J., & Andrews, S. (2012). Lexical quality and reading skill: Bottom-up and top-down contributions to sentence processing. *Scientific Studies of Reading, 16*, 240–262.

Hiebert, E. H. (2012). What's important to know about words of written language. In M. Hougen & S. Smartt (Eds.), *Fundamentals of literacy instruction and assessment, Pre-K-6* (pp. 138–149). Baltimore, MD: Paul Brookes Publishing.

Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. *Explanatory item response models* (pp. 189–212). New York: Springer.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437–447.

Kearns, D. M. (2015). How elementary-age children read polysyllabic polymorphemic words. *Journal of Educational Psychology, 107*, 364–390.

Kearns, D. M., & Al Ghanem, R. (2014). Critical factors in polysyllabic word reading: Roles of letters, sounds, and meaning. Paper presented at the annual conference of the council for exceptional children, Philadelphia, PA.

Kearns, D. M., Steacy, L. M., Compton, D. L., Gilbert, J. K., Goodwin, A., Cho, E., et al. (in press). Modeling polymorphemic word recognition: Exploring differences among children with early-emerging and late-emerging word reading difficulty. Journal of Learning Disabilities.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research, 1*, 1–16.

Kim, Y., Park, C. H., & Park, Y. (2013). Is academic language a separate dimension in beginning writing? Evidence from Korean. *Learning and Individual Differences, 27*, 8–15.

Klauda, S. L., & Guthrie, J. T. (2015). Comparing relations of motivation, engagement, and achievement among struggling and advanced adolescent readers. *Reading and Writing, 28*(2), 239–269.

Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System, 31*, 537–561.

Lee, S. H., & Muncie, J. (2006). From receptive to productive: Improving ESL learner's use of vocabulary in a postreading composition task. *TESOL Quarterly, 40*(2), 295–320.

Mancilla-Martinez, J. (2010). Word meanings matter: Cultivating English vocabulary knowledge in fifth-grade Spanish-speaking language minority learners. *TESOL Quarterly, 44*(4), 669–699.

Maylath, B. (1996). Words make a difference: The effects of Greco-Latinate and Anglo-Saxon lexical variation on college writing instructors. *Research in the Teaching of English, 30*(2), 220–247.

McKeown, M. G., Beck, I. L., & Sandora, C. (2012). Direct and rich vocabulary instruction needs to start early. In E. J. Kame'enui & J. F. Baumann (Eds.), *Vocabulary instruction: Research to practice* (2nd ed., pp. 17–33). New York: Guilford.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*, 237–270.

Nagy, W. E., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47*, 91–108.

National Commission on Writing. (2004). *Writing: A ticket to work…or a ticket out: A survey of business leaders.* Retrieved June 24, 2009, from http://www.writingcommission.org/report/html.

National Commission on Writing. (2005). *Writing: A powerful message from state government.* Retrieved June 24, 2009, from http://www.writingcommission.org/report/html.

Neilsen, L., & Piché, G. L. (1981). The influence of headed nominal complexity and lexical choice on teachers' evaluation of writing. *Research in the Teaching of English, 15*(1), 65–73.

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing, 22*, 545–565.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2012). Vocabulary assessment: Making do with what we have while we create the tools we need. In E. J. Kame'enui & J. F. Baumann (Eds.), *Vocabulary instruction: Research to practice* (2nd ed., pp. 231–255). New York, NY: Guilford.

Perfetti, C. (1992). The representation problem in reading acquisition. In P. B. Gough & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

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

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

Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly, 45*, 8–38.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage Publications.

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading, 7*, 219–237.

Richter, T., Isberner, M. B., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading, 17*, 415–434.

Ricketts, J., Nation, K. & Bishop, D. V. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*(3), 235–257.

Ritchey, K. D., Silverman, R. D., Schatschneider, C., & Speece, D. L. (2015). Prediction and stability of reading problems in middle childhood. *Journal of Learning Disabilities, 48*(3), 298–309.

Schreuder, R., & Baayan, R. H. (1995). Modeling morphological processing. In L. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Lawrence Erlbaum.

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition, 55*, 151–218.

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

van den Noortgate, W., de Boeck, P., & Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics, 28*, 369–386.

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology, 22*, 407–423.

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

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53*, 13–40.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60*, 502–529.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duuvuri, R. (1995). *The educator's word frequency guide [CD-ROM].* New York: Touchstone Applied Science Associates.