Theory and Practice of Second Language Acquisition
vol. 9 (2), 2023, pp. 1/31
https://doi.org/10.31261/TAPSLA.13865

K. James Hartshorn
Brigham Young University, Provo, Utah, USA
https://orcid.org/0000-0002-0629-7410

Aylin Surer
Brigham Young University, Provo, Utah, USA
https://orcid.org/0009-0006-5490-8518

Contributions toward Understanding the Acquisition of Eight Aspects of Vocabulary Knowledge

Abstract

With the intent of adding to the literature leading toward a more complete theory of second language vocabulary acquisition, this study elicited accuracy data from 110 ESL learners ranging from novice high to advanced low on 64 words randomly selected in the 2K–3K range of Corpus of Contemporary American English (COCA) (32 verbs, 24 nouns, 8 adjectives) covering eight aspects of word knowledge. These included spelling based on hearing the spoken form, selecting collocations based on the written form, pronunciation based on the written form, selecting inflections based on the written context, selecting the definition based on hearing the spoken form, selecting the written definition based on the written form, selecting appropriate derivations based on the written form, and selecting the written form based on the written definition. ANOVA results show accuracy levels varied across word knowledge aspects and that implicational scaling was possible with some but not all aspects of word knowledge examined simultaneously. In aggregation with other current and future studies, this has important implication for developing L2 vocabulary acquisition theory.

Keywords: second language vocabulary acquisition, aspects of word knowledge, implicational scaling

For four decades, scholars have lamented the lack of a complete theory of second language vocabulary acquisition (e.g., Meara, 1983; Schmitt, 1995, 2019). Nevertheless, some incremental progress has been made. For instance, we now have valuable insights regarding vocabulary coverage needed for text comprehension (Laufer, 1989, 1992; Hu & Nation, 2000; Nation, 2006; Schmitt et al., 2011). Scholars such as Richards (1976) and Schmitt (1998) have also described various components of word knowledge and suggested that some

may be interrelated and that their acquisition may be incremental (Schmitt, 1998). More recently, González-Fernández and Schmitt (2020) have used implicational scaling to suggest an acquisition order for a number of aspects of word knowledge. This progress is promising for increasing vital insights about second language vocabulary acquisition. Nevertheless, more complementary and confirmatory data are needed from multiple streams of evidence across many contexts if we are to solidify our knowledge of how vocabulary is acquired and whether a durable acquisition order for various word knowledge components can be established (e.g., González-Fernández & Schmitt, 2020; Schmitt, 2019). Such insights would be invaluable for L2 teachers, materials developers, theorists, and researchers alike. Therefore, this study was designed to provide important contributions to the literature by identifying an accuracy order for second language learners on eight specific aspects of word knowledge in an ESL context.

Review of Literature

A robust knowledge of vocabulary is fundamental to second language development and comprehension. In reading, for example, many scholars agree that comprehension requires mastery of approximately 95 to 98% of the words readers encounter (Laufer, 1989; Hu & Nation, 2000; Schmitt et al., 2011). Thus, vocabulary acquisition is an essential component in language development. Though at its most fundamental level, vocabulary acquisition requires knowledge of a word's "form" and "meaning" (Thornbury, 2002, p. 15), much more can be included in what it means to know a word. For example, Richards (1976), described word knowledge as including an understanding of the word's form, meaning, frequency, syntactic features, derivations, associations, and the various limitations on the use of the word.

In an effort to describe word knowledge, some researchers have examined vocabulary development in terms of breadth and depth (Chapelle, 1998; Qian & Schedl, 2004; Schmitt, 2014). The notion of word breadth or the number of words known is well correlated with efficacy in writing (Milton et al., 2010; Stæhr, 2008), and speaking (Zimmerman, 2004), as well as in higher levels of comprehension in listening (Stæhr, 2008; Zimmerman, 2004) and reading (Laufer, 1992; Qian, 1999; Stæhr, 2008).

Despite benefits associated with vocabulary breadth, determining the depth of one's vocabulary knowledge seems to be more difficult. The development of various instruments has been useful such as Wesche and Paribakht's (1996) Vocabulary Knowledge Scale, which identifies word familiarity by measuring

vocabulary recognition and production. Another helpful resource has been Read's (1998) Word Associates Format test which examines knowledge of paradigmatic and syntagmatic word associations (Zhang & Koda, 2017). In addition, research has examined the positive effects of vocabulary depth on various skills such as speaking (Koizumi, 2005; Kilic, 2019), listening (Farvardin & Valipouri, 2017; Teng, 2014), writing (Atai & Dabbagh, 2010; Kilic, 2019), and reading comprehension (Farvardin & Koosha, 2011; Mehrpour et al., 2011; Qian, 1999). Such studies highlight the importance of learners developing both vocabulary breadth and depth.

Nation (2001) has also suggested that a more complete understanding of vocabulary depth is needed. He described word knowledge as a word's form (including the spoken form, written form, and word parts), meaning (including connections between form and meaning, concepts and referents, and associations), and use (including grammatical functions, collocations, and various constrains on the use of a word). Thus, word knowledge could refer to an individual's facility with each of these nine elements. Yet, because aspects of word knowledge can be examined productively and receptively, Nation's nine components could be expanded to eighteen.

Despite these numerous aspects of word knowledge, however, specific attempts to operationalize data elicitation of word knowledge could further expand the number of contexts worth studying. For example, consider the various types of stimuli that might be used to prompt a learner to write a specific word. In the L2, learners might hear the word, one or more definitions, a derivation, or an inflection. Or, they might read a definition, a synonym, an antonym, a derivation, an inflection, and so forth. Conversely, they might encounter these or many other types of prompts in their L1. Alternatively, prompts may be much less direct, or language data may be based on completely natural production with no prompt at all. Although the specific task for the learner to write a particular word may be the same across settings, performance levels may vary widely depending on the exact nature of the stimuli, the context, and the learners themselves. This variability should be taken into account in vocabulary acquisition studies.

Relationships among Word Knowledge Components

González-Fernández and Schmitt (2020) have noted that while most studies currently available have examined only one aspect of word knowledge at a time, this approach may be inadequate for developing a more complete understanding of vocabulary acquisition. Rather they "encourage the measurement of multiple components concurrently" (p. 483). A few studies have simultaneously examined a small number of word-knowledge components. For

example, in their research on the effects of lexical depth and breadth on reading comprehension, Qian (2002) examined synonymy, polysemy, and collocations. Pellicer-Sanchez and Schmitt (2010) examined word class, word recognition, spelling, and recall of meaning. Wesche and Paribakht (1996) developed the Vocabulary Knowledge Scale and had students describe their level of word knowledge in terms of word production and recognition.

Several scholars have undertaken studies designed to reveal key relationships among various aspects of word knowledge. For example, over the course of one year, Schmitt (1998) examined the development of four of these, including senses of meaning, spelling, associations, and grammatical features. Schmitt concluded that some of these aspects of word knowledge seemed to be related in their development. He noted that senses of meaning were more closely related to grammatical features and associations than grammar. Schmitt also observed that spelling was generally acquired before the other aspects of word knowledge. Despite these insights, Schmitt was unable to identify a valid implicational scale showing a developmental hierarchy across word knowledge components due to inconsistencies in his data.

Looking at both receptive and productive contexts, Webb (2005) examined five aspects of word knowledge including meaning, grammatical features, syntax, association, and orthography. He observed that strategies associated with productive skills generated more productive and receptive knowledge of orthography, meaning, syntax, and grammar but the strategies associated with receptive learning only produced more receptive knowledge of meaning. He advocated the use of instruments that measure both productive and receptive word-knowledge components. Later, when examining the effects of repetition, Webb (2007) noted that some aspects of word knowledge emerged before others. For example, receptive knowledge syntax, grammatical features, orthography, and productive knowledge of association emerged before meaning. Building on the work of Webb (2005, 2007), Chen and Truscott (2010) similarly observed language development for both receptive and productive aspects of word knowledge for orthography, part of speech, associations, and meaning and form, though they noted that the link between form and meaning took longer to be mastered compared to the other components.

Laufer and Goldstein (2004) examined four aspects of word knowledge including active recall where the learner produces the target word), passive recall where the learner provides the word's meaning, active recognition where the learner identifies the word from a list of options, and passive recognition where the learner identifies the meaning of the word from a list that includes distractors. As they hypothesized, accuracy levels for these tasks showed a clear difficulty order ranging from easiest to most difficult: passive recognition, active recognition, passive recall, and finally active recall.

With an emphasis on the acquisition order of various aspects of word knowledge, González-Fernández and Schmitt (2020) identified a valid implicational scale based on difficulty for these components in writing. From most accurate to least accurate, these include: (a) form-meaning recognition, (b) collocate form recognition, (c) multiple meaning recognition, (d) derivative form recognition, (e) collocate form recall, (f) form-meaning recall, (g) derivative form recall, and (h) multiple meaning recall. They concluded that the form-meaning link is more difficult than productive and receptive knowledge of orthography, part of speech, and associations. They also suggested that the form-meaning link is easier for learners to master than collocations, multiple meanings, and derivatives.

Though few studies have examined multiple aspects of word knowledge simultaneously, the work of González-Fernández and Schmitt (2020) provides important new insights with the generation of a valid implicational scale. Though cross-sectional rather than longitudinal, these findings suggest incremental development and a hierarchical order of the various aspects of word knowledge examined. Nevertheless, González-Fernández and Schmitt (2020) acknowledge that the construct of vocabulary knowledge is based on many more aspects than can possibly be examined effectively in one study and that many more studies are needed. They have suggested that "future studies should explore different combinations of components to build a composite picture of the overall word knowledge component constellation" (p. 501).

Scholars interested in answering this call to contribute should also consider the many valuable suggestions regarding this line of inquiry. One challenge has to do with "test contamination […] where exposure to a target word on one test […] may give hints to answering a subsequent test" (Schmitt, 2019, p. 263). A potential solution could be to utilize different words in different instruments rather than the same set of words across aspects of word knowledge (e.g., Kieffer & Lesaux 2012; Li & Kirby 2015; Milton & Hopkins 2006). González-Fernández and Schmitt (2020), who only used twenty words in their study, recommended that researchers use a larger sample of words and that researchers should include students from heterogenous L1 backgrounds rather than a single L1 background.

Based on the preceding review, and in consideration of these important suggestions, the current study was designed to add to the literature by examining a complementary set of aspects of word knowledge. The aspects selected for this study were based on the literature as well as constraints inherent to our research context and include some of the most common tasks associated with what it means to know a word.

Research Questions

As mentioned previously, testing various aspects of word knowledge in this study are operationalized as particular tasks based on specific prompts. These include: (a) spelling based on hearing the spoken form, (b) selecting collocations based on the written form, (c) pronunciation based on the written form, (d) selecting inflections based on the written context, (e) selecting the definition based on hearing the spoken form, (f) selecting the written definition based on the written form, (g) selecting appropriate derivations based on the written form, (h) selecting the written form based on the written definition. With these targeted aspects of word knowledge in mind, the following research questions are articulated:

1. To what extent does the accuracy of ESL learner performance vary across the specified eight aspects of word knowledge?
2. Do accuracy levels of ESL learner performance across the specified aspects of word knowledge form an implicational scale?

Methods

This section describes the selection of the words used in this study, the development of the instrument, the learners who provided data for this study, and the planned analyses.

Word Selection

Building on the recommendation of González-Fernández and Schmitt (2020) to use more than twenty words, a total of 64 words were selected to represent eight different aspects of word knowledge. These words were initially chosen randomly from between frequency rankings of 2K–3K in the Corpus of Contemporary American English (Davies, 2008). This frequency range was selected based on previous assessments that suggested that many of these words would be known by the advanced proficiency learners but not by the novice learners. It was expected that such a range in word knowledge would be necessary for implicational scaling. It was intended that a representative list of words from different parts of speech be used that could help answer the research questions associated with the different aspects of word knowledge of interest in this study. Some adjustments from the original randomized list were made to ensure that all words could have derivational and inflectional forms.

Adverbs were not used in this study since they do not undergo inflection in English. The final list included 32 verbs, 24 nouns, and eight adjectives (see Appendix A for the complete list).

Instrument Development

This section details the creation of the data instrument used in this study. The instrument was developed as an electronic survey to be delivered to student email addresses during a class period in the IEP's computer lab with monitoring provided by the students' teachers and the researchers. As described above, the instrument was designed to test eight different aspects of word knowledge, using eight words to establish mastery for each aspect. Each of these item types will be described below.

At the outset, however, we begin with a brief description of the creation of the audio recordings used in this study. Two of the item types in the instrument required audio recordings of the words of interest. Audio recordings to be included were made using Adobe Audition CC 2019 and the built-in microphone in a 2019 MacBook Pro with the speaker's voice one and a half feet away from the microphone. Postproduction included reduction of ambient noise using the default setting of the DeNoise effect. Each recording was also normalized to 95%. Minor post-production editing resulted in the final recordings for each word beginning with 500 milliseconds of silence followed by a first audio presentation of the word of interest. This was followed by two seconds of silence and then a second production of the word. This was done for all 64 words and example recordings used to introduce item types that utilized audio. We now provide a brief description of each item type.

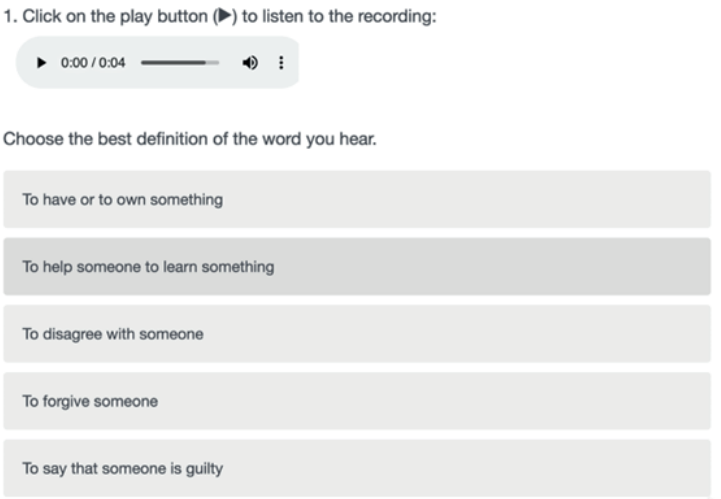
Recognizing the Meaning from Hearing the Word

The first item type in the instrument provided students with the audio and then invited them to choose the best definition of the word they heard by using their mouse to select the most appropriate response. Figure 1 illustrates this item type for the word "accuse." Students clicked on the play button to hear the audio and then selected the best definition from among five options. Distractor definitions were randomly selected from other words within the 2K–3K range. In the very few cases where the randomly selected definition shared a meaning sense with the target word, another definition was randomly chosen so there would be only one correct response. As shown in the figure, definitions were kept short and utilized high frequency vocabulary. This was done with the intent that incorrect responses would be based on not knowing the meaning of the word rather than challenges associated with reading or understanding the

options within the item. Each of the eight items of this type were simply scored as correct (1 point) or incorrect (0 points) depending on the answer.

Figure 1

Sample Item for Recognizing a Written Definition Based on the Spoken Form

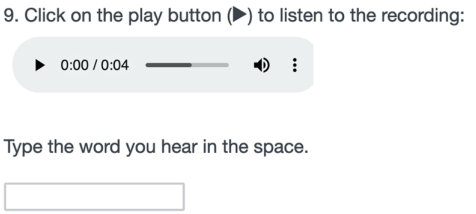


Spelling the Word

The second aspect of word knowledge tested student ability to spell a word based on hearing the word. The prompt for this item type was the same as the previous item in that students were presented with the audio in the same format. After clicking on the play button to initiate the audio, students were invited to type the word in a provided textbox as illustrated in Figure 2. Scoring was limited to the actual spelling of words without regard to capitalization. No attempt was made to give partial credit. This item type was scored with 1 point for each correctly spelled word, and no points for any misspelled words.

Figure 2

Sample Item for Spelling the Word based on the Spoken Form

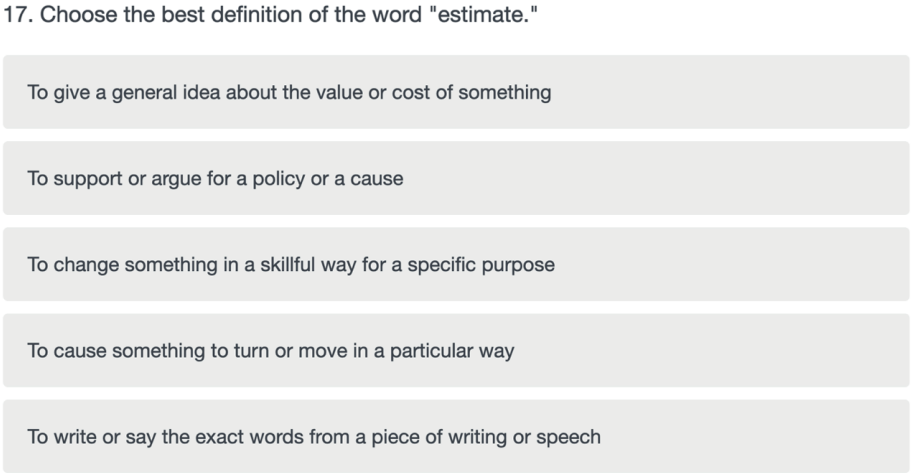


Recognizing the Meaning from the Written Form

The third item type in the elicitation instrument presented students with the written form of the word and then invited them to identify the best written definition of five possible options. As with previous items, definitions were randomly selected. They were also kept relatively short and utilized higher frequency vocabulary than the word being defined. Figure 3 provides a sample of this item type from the instrument. As with previous items, students were given one point for each correct answer and no points for any wrong answers.

Figure 3

Sample Item for Selecting the Definition Based on the Written Form

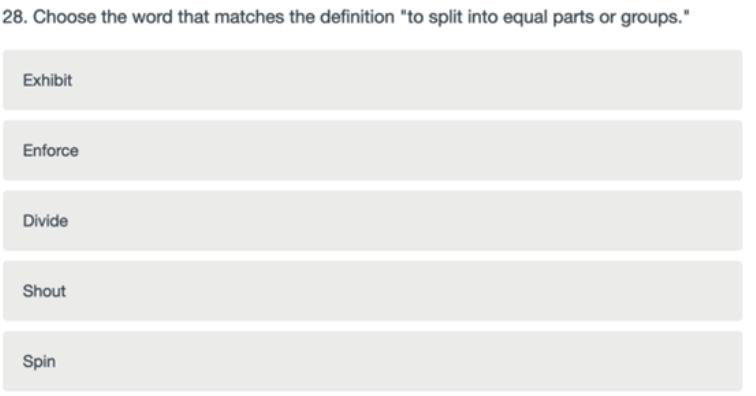


Recognizing the Word from a Written Definition

The fourth item type in the instrument was the inverse of the previous item. Students were presented with a simple definition and invited to select the word that was the best match for the definition. As with previous item types, words were randomly selected from within the 2K–3K frequency band. Figure 4 provides an example of this item type. As with previous tasks, students were awarded one point for each correct response and no points for incorrect answers.

Figure 4

Sample Item for Selecting the Written Word Based on the Written Definition

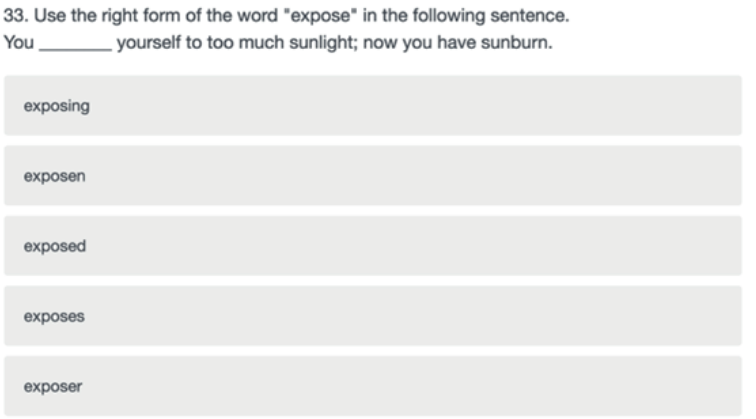


Recognizing Appropriate Inflections

The fifth item type was designed to test learner knowledge of word inflection. Students were provided with the uninflected word of interest and a sentence requiring an inflected form of the word. Students chose from among five options. Distractors were formed by adding inflectional morphemes common in English but that were not appropriate for the context. Figure 5 illustrates this item type from the instrument for the word "expose." Correct answers were given one point and incorrect answers with given no points.

Figure 5

Sample Item for Selecting the Written Word Based on the Written Definition

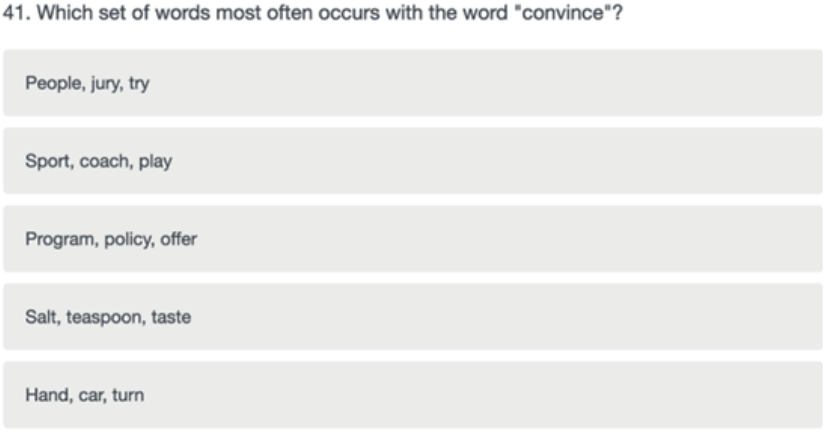


Collocations

The next item type was designed to test learner knowledge of collocations for each word of interest. The collocations used in our instrument were based on information provided in the frequency dictionary by Davies and Gardner (2010). Our intent was to choose three of the most common collocations for each word included in the instrument. Figure 6 provides a sample item from the instrument where "people," "jury," and "try" are common collocations for the word "convince" (Davies & Gardner, 2010, p. 121). We note that "people" and "jury" were the first two collocations listed in the dictionary under the noun category and that "try" was the first entry under the miscellaneous category. Though this entry for "convince" only included a noun and miscellaneous categories, other entries included additional categories. For example, the word "pepper" in the dictionary includes the adjectives ("red, black, green, hot…"), nouns ("salt, teaspoon, bell…"), and verbs (taste, add, chop, dice…"). In such cases, we generally went with the first word from each category such that the correct response for "pepper" would be "red, salt, taste." Responses were scored with one point for correct answers and no points for incorrect answers.

Figure 6

Sample Item for Selecting Collocations Based on the Written Form



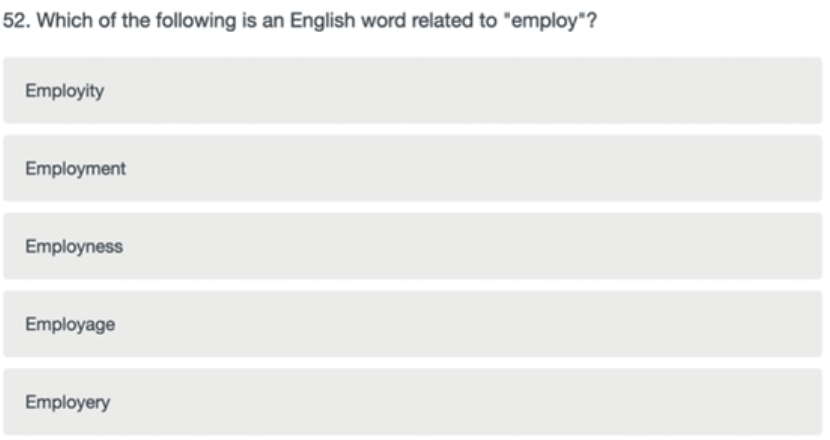
Derivations

This item type was designed to test learner knowledge of derivations of the target words emphasized in the instrument. Learners were presented with the word and then invited to choose which of the five options was an actual word in English based on the written form of the word of interest. Distractors were

generated by using nonwords that were morphologically related to the word and were designed to appear as the same part of speech. Figure 7 illustrates this item type for the word "employ." As with other items, one point was given for each correct response.

Figure 7

Sample Item for Selecting an Appropriate Derivation Based on the Written Form



Oral Production Based on Written Form

The final item type was designed to test the learner's ability to appropriately pronounce the word in context. In this case, one of the target words was situated in one of eight sentences presented to the learner for them to read aloud while being recorded. As with previous items, care was given to keep the sentences relatively short and to ensure that the other included words were of higher frequency than the word of interest. Though short, complete sentences were used to help differentiate polysemous forms such as "suspect" (in the second sentence below) which could be interpreted as a noun or a verb with different phonological forms without the context provided by the sentence. The software used to record learner voices was proprietary and had been installed on the computers in the lab where data were collected. One point was given for each correctly pronounced word. However, scoring for this item type was more complicated due to the need to establish inter-rater reliability estimates which will subsequently be described in more detail. This item type is displayed in Figure 8.

Figure 8

Sample Item for Oral Production of a Word Based on the Written Form

Say each of these sentences:

1. I pray every day.
2. I suspect he is the thief.
3. I love reading that column.
4. He doesn't oppose them.
5. Nobody solved the mystery.
6. We will transform our kitchen.
7. It is a rare bird.
8. He is a champion.

As described previously, the elicitation instrument included eight words for each of the eight item types for a total of 64 words. However, since the questions for certain aspects of word knowledge could give away the answers for other aspects of word knowledge, the instrument had to be carefully constructed. For example, hearing the spoken form of the word for one item testing one component of word knowledge could alert the learners how to pronounce the same word for an item testing a different component of word knowledge. To avoid this problem, eight different test forms were created. This allowed testing that included all 64 words used for the eight different components of word knowledge but that did not use the same words across forms to elicit data on the same aspect.

For example, consider Table 1 that illustrates the distribution of just eight words (represented by letters A–H) across the eight different aspects and eight different test forms. Let's say the letter "A" represents the word "accuse." In Form 1 of the instrument, the word "accuse" is used to test the first aspect of word knowledge. Therefore, the student hears the word "accuse" and selects the best definition. In Form 2 of the instrument, however, the word "accuse" is used to test the second aspect of word knowledge. So, the student hears the word and types the word "accuse" in the space provided. Thus, in summary, all students were tested on all eight aspects of word knowledge using the same 64 words, though not all students were presented with the same words for the same aspects across the eight different test forms.

Table 1

Distribution of Words Across Aspects and Test Forms

	Test Forms							
Aspect	1	2	3	4	5	6	7	8
1	**A**	H	G	F	E	D	C	B
2	B	**A**	H	G	F	E	D	C
3	C	B	**A**	H	G	F	E	D
4	D	C	B	**A**	H	G	F	E
5	E	D	C	B	**A**	H	G	F
6	F	E	D	C	B	**A**	H	G
7	G	F	E	D	C	B	**A**	H
8	H	G	F	E	D	C	B	**A**

Though we acknowledge this is an imperfect data elicitation solution since the respective forms are not exactly the same, we believed that in aggregate, this approach would prevent the elicitation instrument from inappropriately revealing additional word information to the participants. We also believed that the potential benefits associated with new insights from this strategy likely outweighed the potential limitations of this approach. Also, since it is conceivable that the ordering of particular aspects of word knowledge could impact learner performance, the order of these different item types were presented randomly within the different forms of the elicitation instrument.

Participants

This study was sponsored by the intensive English program where the study occurred with the express intent that results could help inform materials development and pedagogy. Accordingly, all ethics standards were met in the gathering of these data. Unfortunately, due to the COVID-19 pandemic, student enrollment in the program was less than half its typical number. Thus, only 110 students provided data for all eight of the aspects of word knowledge examine in this study. Of the participating students, there were 58 females and 52 males. Participant ages ranged from 18 to 57 though most students were in their twenties (M = 24.37; SD = 6.29). Although about two-thirds of the students were native speakers of Spanish (73), other L1s included Japanese (16), Portuguese (6), Chinese (4), French (3), Haitian Creole (3), Korean (2), Russian (2), and Albanian (1). Proficiency levels ranged from novice high to advanced low according to guidelines from ACTFL (*American Council on the Teaching of Foreign Languages*, 2012) as illustrated in Table 2.

Table 2

Proficiency Levels of Student Participants

Proficiency	*N*	%
Advanced Low	20	18.18
Intermediate High	30	27.27
Intermediate Mid	40	36.36
Intermediate Low	14	12.73
Novice High	6	5.45
Total	110	100.00

Raters

Though most data examined in this study did not require a reliability estimate, the interrater reliability for oral production of the words was established by the two authors, one of whom holds a Ph.D. and who has worked in the field of second language teaching and learning for more than three decades. The other holds an MA in TESOL and has taught EFL/ESL for about seven years.

Analyses

Interrater reliability for oral production of the words examined in this study was established by the authors based on two broad categories. The first was the phonological appropriateness of the production, and the second was the appropriate stress accent based on the word in a simple sentence. Though the raters agreed that some latitude would be allowed for slight departures from phonological norms, any overtly conspicuous phonological substitution of consonants or vowels would be considered an error. Similarly, any obvious departures from stress accent norms would also be considered an error. Raters only evaluated the specific words targeted for the study, so any additional departures from pronunciation norms within the sentences were ignored. Rating involved evaluating each of the eight words used to test oral production, resulting in a rating deemed correct (1) or incorrect (0) for each word. Thus, raters provided each student with a raw score ranging from 0 to 8. In six cases, recordings for one or more of the words were unexpectedly cut short. Rather than completely discard data from these students, the missing data were replaced with mean performance levels for the items for which recordings were available.

While one researcher provided a rating for each student included in the study, the other randomly rated 70% of the group. This initially produced a Pearson correlation of .85 ($p < .001$). However, examination of the data revealed four

cases with a rating difference of two or more. Without discussing any details about these cases, raters were invited to reexamine these recordings to ensure no clerical mistakes or other oversights had produced the discrepant scores in error. After reexamination, some corrections to these cases were made, with a resulting Pearson correlation of .92 ($p < .001$). In an effort to leverage the perceptions from both raters, averages were calculated for those students with two ratings. Subsequent analyses were based on these scores.

The intent was that each test form functions similarly for each aspect of word knowledge. Test forms were randomly assigned within each proficiency level. Though not all students completed the test, the number of students taking each test form and their respective proficiency levels were fairly well distributed, with no statistically significant difference across performance levels for the test forms themselves, $F(7,102) = .089$, $p <. 999$ (see Table 3).

Table 3

Descriptive Statistics and Proficiency Level by Test Form

	Proficiency Level					Descriptives		
Form	NH	IL	IM	IH	AL	*N*	*M*	*SD*
1	1	2	6	4	2	15	335	110
2	1	2	5	4	2	14	337	114
3	0	3	4	4	2	13	344	104
4	0	3	4	4	3	14	355	108
5	1	1	4	3	3	12	356	123
6	1	1	6	4	2	14	344	107
7	1	1	6	3	3	14	351	114
8	1	1	5	4	3	14	358	114
Total	6	14	40	30	20	110		

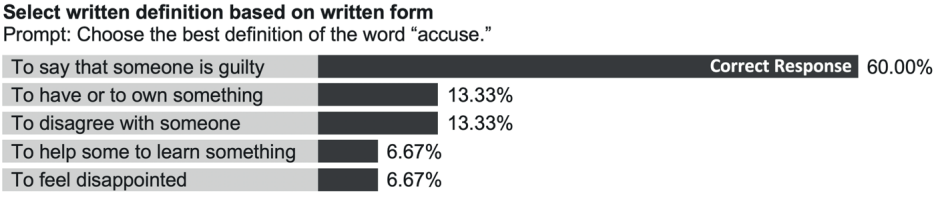
Moreover, no significant difference was observed for performance levels across test forms for five of the eight aspects of word knowledge including spelling based on hearing the spoken form, $F(7,102) = 0.49$, $p = .84$, selecting the inflection based on the written context, $F(7,102) = .519$, $p = .819$, pronouncing the word based on the written form, $F(7,102) = .577$, $p = .773$, selecting the definition based on hearing the spoken form, $F(7,102) = .667$, $p = .691$, and selecting the derivation based on the written form, $F(7,102) = 1.708$, $p = .115$. However, the original performance levels for three of the aspects of word knowledge were not uniform across test forms including selecting the written form based on the written definition, $F(7,102) = 2.307$, $p = .032$, selecting the collocations based on the written form of the word, $F(7,102) = 3.726$, $p = .001$, and selecting the written definition based on the written form of the word, $F(7,102) = 4.945$, $p < .001$. Given these discrepancies, test form effect was accounted for and eliminated in subsequent analyses.

Implicational scaling was used to address the second research question (e.g., Hakansson, 2013; Hatch & Lazaraton, 1991; Rickford, 2002). Implicational scaling can be used to show which aspects of word knowledge may be the easiest or most difficult for learners to master. If an accuracy order is scalable, it may suggest an acquisition order. Implicational scaling has been widely used for hierarchical ordering of "grammatical, lexical, and phonological features of language" (Hatch & Lazaraton, 1991, p. 204) with important applications for teaching and second language materials development. For this study, the accuracy threshold for each aspect of word knowledge was set at 75%. Though this threshold is on the lower end of the acceptable range, typically between 75–90% accuracy (e.g., Dulay & Burt, 1974; Ellis, 1988), this level was chosen with the hope it might help mute error levels that might be introduced by using different test forms. Since the use of longitudinal data were not feasible for this study, implicational scaling was based on cross-sectional accuracy data gathered on a single occasion.

Before presenting the findings designed to answer our research questions, we briefly examine student responses in greater detail. Items used for elicitation were of three types. These include several multiple-choice formats as well as the spelling and spoken production of specified words. Figure 9 shows the distribution of responses for one multiple choice item type seeking the best definition of the word *accuse*. This illustrates the typical pattern with most students responding correctly while others chose various distractors. For additional examples of multiple-choice responses, see Appendix B.

Figure 9

Example of Response Distributions



Perhaps more informative than the multiple-choice items, however, are the variety of productive responses of spelling and pronunciation. Though extensive analysis of these errors is beyond the scope of this work, a few examples and comments about errors may be useful. In spoken production, some errors were phonologically similar English words though not those elicited such as *poor* for *pure*, *pose* for *oppose*, *rear* for *rare*, *pry* for *pray*, *concrete* and *complete* for *compete*, and so forth. In some cases, students substituted one or more erred phonemes such as /ˈbæʃən/ for *passion*, /tɹænzˈfɔrn/ for *transform*, and /ˈtʃɑɹp/

for *sharp*. In other cases, students altered or omitted one or more phoneme such as /səsˈpɛt/ for *suspect*, /ˈmɪsɪri/ for *mystery*, and /kɑnˈvaɪz/ for *convince*, and /ˈæksə/ for *access*. Still in other cases, productions shared only vague similarities with the elicited words such as /ˈfiɛt/ for *thief*.

In terms of spelling, just four of the sixty-four words included in the study were spelled correctly by all participants including *mix*, *invest*, *suspect*, and *emotion*. Words which generated five or more misspellings are presented alphabetically in Table 4. Similar to some pronunciation errors, some words or phrases were spelled correctly but were not those elicited by the prompts. These include errors such as *a quarter* for *acquire*, *quiz* for *accuse*, *uplift* for *athlete*, *belief* and *breath* for *brief*, *complete* for *compete*, *device* for *divide*, *ask to me* for *estimate*, *mistreat* for *mystery*, *vacation* for *occasion*, *orange* for *origin*, *poor* for *pure*, and *strait* and *stretch* for *straight*. Possibly due to limitations in working memory, some students also produced errors by inappropriately inflecting target words such as *attracted* for *attract*, *employed* for *employ*, and *opposed* for *oppose*.

Other error types seem consistent with predictions from the orthographic depth hypothesis which suggests greater difficulties where orthographies such as English are not well aligned with phonology (Frost, 2005). Many students attempted to use a single letter to represent phonemes spelled with two letters in English. These include misspellings such as *acomplish* for *accomplish*, *acuse* for *accuse*, *aprove* for *approve*, *atract* for *attract*, *colum* for *column*, *ocasion* for *occasion*, *pasion* for *passion*, and so forth. Similarly, other mistakes may have been associated with multiple letters or word formation patterns in English that represent the same or similar sounds, resulting in constructions such as *mistery* for *mystery*, *filozofi* and *phylosofy* for *philosophy*, *strait* and *strate* for *straight*, *breaf* and *breef* for *brief*, *welth* for *wealth*, and so on. Additional research may be needed to better understand these spelling error patterns more fully.

Table 4

Words with Five or More Misspellings

Word	Misspellings
Accomplish	acmplish, acomplesh (2), acomplish (5), acoplish
Accuse	acuse (9), quiz
Acquire	a quarter, aquair, aquaire (2), aquare, aquareir, aquarer
Approve	aprofe, aprouve, aprove (6)
Athlete	afflide, aflate, afleed (2), aflict, aflied, aflix, afraid, afread, afrid (2), uplift
Attract	atrack, atracked, atract (3), attracted, atractt, attarct
Brief	belief, breaf (2), breath, breef, brive (2)
Column	calam, colam, colom (3), colon (3), colonne, colum, coron

Compete	compet, competa, compite (6), complete
Convince	convence (3), convens, convience
Divide	devaed, device, devid, devide (5), duvaret
Emphasize	emphazise, emphese, emphsize, enfacides, enfasis, enphase, inphasize
Employ	employe (2), employed (3), imploe, impory, improal
Estimate	ask to me, astomate, attrac, estmate, estmit, stament, stimate
Fiction	ficcion, ficion, ficttion, fitshen, fixion
Mystery	mistry, mistery (6), mistread, mistreat
Occasion	acation, ackigan, ocasion (2), ocassion, ocation (3), occation, vacation
Oppose	apos, apous, appose, appouse, opositive, oposive, opouse, opposed
Origin	orange (2), orgen, origaine, origen (3), origine
Passion	pacient, pacient, partsion, pasion (2), pation
Permit	permet, permite, premitted, promed, promet (2)
Philosophy	filosophy (2), filozofi, forasefi, forasefy, phirosify, phylosofy, phylosophy
Pure	pior poor, priort, puler, puor, pur
Scholar	schollar, scholor, schoolar (3), schooler, (3) scoger, scolar, skoler
Smooth	slud, smode, smooded, smoose, smoth
Straight	schoolar, straght, strait (2), straith, strate, streat, streid, strenge, stretch
Symbol	sambal, sembal, simbol, simbole (3), symbole
Wealth	walth, weld, welf (3), welft, welth (2)

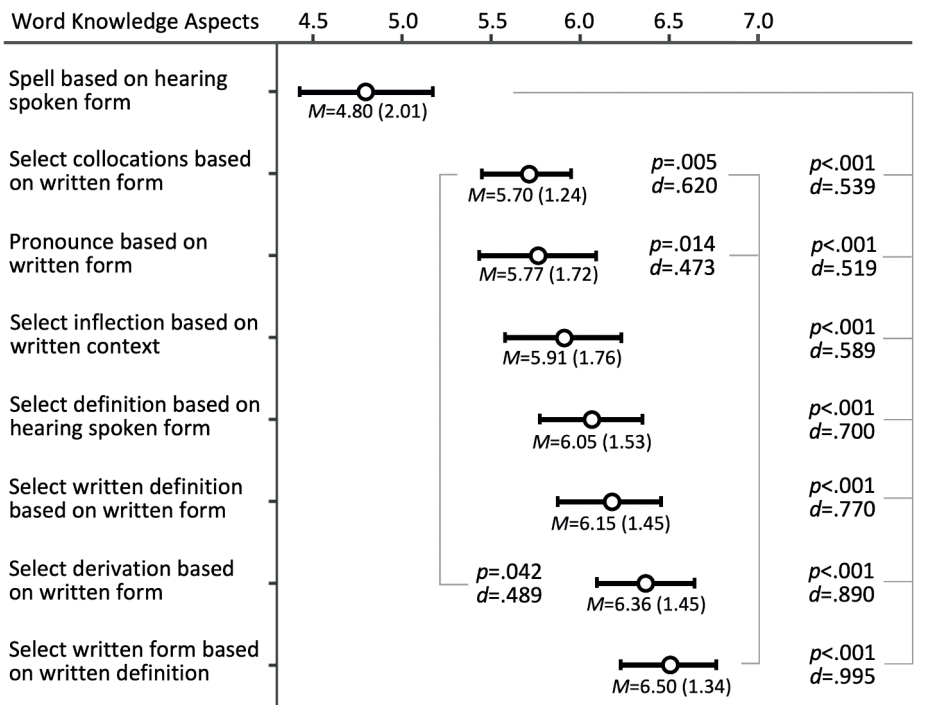
Note: Parentheticals indicate the number of observations of the same spelling.

Results

This section presents findings associated with the two research questions. The first question addressed the extent to which the accuracy of ESL learner performance varied across the eight aspects of word knowledge examined in this study. Results of a one-way ANOVA indicated that performance levels indeed varied across aspects of word knowledge, $F(7,872) = 12.1$, $p < .001$, and a Tukey post-hoc test showed statistically significant differences between specific aspects of word knowledge. Figure 10 illustrates these differences, presenting means, standard deviations (in parentheses), p-values, and effect sizes. Performance on these aspects of word knowledge were based on a possible range of 0 to 8 and are arranged from least accurate at the top of the figure to most accurate at the bottom of the figure.

Figure 10

Mean Accuracy Levels across Aspects of Word Knowledge



The second research question addressed whether ESL learner performance across aspects of word knowledge form an implicational scale such that mastery of one aspect would suggest mastery of one or more other aspects. Though the findings illustrated in Figure 10 provide general evidence that the ESL performance levels varied across components of word knowledge, an implicational scale could not be formed utilizing all eight components of word knowledge simultaneously. Nevertheless, implicational scaling was successful with some subsets of the total list.

For instance, implicational scaling was achieved[1] with the following aspects of word knowledge: being able to spell a word based on hearing the spoken form ⊂ being able to pronounce the word based on the written form ⊂ being able to recognize the written from based on the written definition. In other words, these data suggest that accurate spelling implies the ability to pronounce the word, and accurate pronunciation implies the ability to identify the written form based on the definition (Crep = .927; MMrep = .333; %Imp =.5937; Cscal = .89).

[1] In order to claim scalability, usually the coeffect of reproducibility (Crep) must be ≥ .90 and the coefficient of scalability (Cscal) must be ≥ .60 (Guttman, 1944).

Similarly, the following slightly varied list was also scalable: being able to spell a word based on hearing the spoken form ⊂ being able to select appropriate inflections based on the written form of the word ⊂ being able to recognize the written form based on the written definition (Crep = .939; MMrep = .333; %Imp =.606; Cscal = .909).

Other potential scales were also observed though they merely approached but did not meet the traditional expectation of .90 for the coefficient of reproducibility. Here are three of these. The first includes being able to spell a word based on hearing the spoken form ⊂ being able to pronounce the word based on the written form ⊂ being able to select a definition of the word based on hearing it ⊂ being able to recognize a definition of the word based on reading the written form ⊂ being able to recognize the written form based on the written definition (Crep = .866; MMrep = .351; %Imp =.515; Cscal = .793). The second includes being able to spell a word based on hearing the spoken form ⊂ being able to select an appropriate inflection based on the written form ⊂ being able to select an appropriate derivation based on the written form ⊂ being able to select an appropriate definition based on reading the word ⊂ being able to select the written form based on the written definition (Crep = .85; MMrep = .333; %Imp =.594; Cscal = .89). The third includes being able to spell a word based on hearing the spoken form ⊂ being able select an appropriate inflection based on the written form ⊂ being able to select a definition based on reading the word ⊂ being able to select the written form based on the written definition (Crep = .886; MMrep = .334; %Imp =.555; Cscal = .83).

Discussion

Building on the work of other scholars including González-Fernández and Schmitt (2020) and employing some innovations in data elicitation, this study sought to examine the extent to which accuracy of ESL learner performance varied across eight aspects of word knowledge and whether ESL learner performance levels would form an implicational scale. Data for this study were elicited through the presentation of certain tasks based on specific types of prompts or stimuli. Though a valid implicational scale could not be formed for all eight aspects of word knowledge examined simultaneously in this cross-sectional study, analysis of variance and implicational scaling of subsets of the complete list of aspects of word knowledge revealed meaningful differences in accuracy levels across components of word knowledge. Thus, these findings may be useful in aggregate with other current and future studies in providing important insight about vocabulary acquisition.

For the eight aspects of word knowledge included in this study, spelling—based on hearing the word—proved to be the most difficult for learners on average. The accuracy levels for spelling were significantly lower than every other aspect of word knowledge observed in this study. The next most difficult feature for learners in this study after spelling was knowledge of collocations. Collocations were significantly more difficult for learners compared to selecting a derivation based on the written form or selecting the written form based on the written definition. The third most difficult aspect of word knowledge was pronunciation of the word based on the written form, which was significantly less accurate compared to selecting the written form based on the definition. Thus, the two item types requiring demonstration of productive skill ended up in the cluster of the three most difficult aspects of word knowledge. Despite clear differences in learner performance levels for these components of word knowledge illustrated in Figure 10, no other differences were observed in performance levels across the other aspects of word knowledge.

At a broad level, such findings showing varied performance levels across aspects of word knowledge are consistent with the studies of other researchers such as Laufer and Goldstein (2004), Webb (2005, 2007), and González-Fernández and Schmitt (2020). Generally, the pattern observed in this study showed that active recall was more difficult than passive recognition consistent with Laufer and Goldstein (2004) and González-Fernández and Schmitt (2020) and that demonstrations of productive knowledge was more difficult than receptive knowledge consistent with Webb (2005, 2007).

Thus, whilst general observations in this study associated with productive and receptive knowledge, recall and recognition, seem consistent with previous research, some discrepancies remain that warrant further study. Noting inconsistencies in previous research, González-Fernández and Schmitt (2020) question "whether all the recall aspects are more difficult than all recognition aspects," as they observed in their study, "or whether some recall aspects can be easier than some recognition aspects" (p. 497). In their earlier research, Pigada and Schmitt (2006) observed that students performed more accurately on the recall component of spelling than they did with the recognition component of grammar knowledge. However, Pellicer-Sanchez and Schmitt (2010) observed that the recall components of word class and meaning were more difficult compared to the recognition components of meaning and spelling.

Though it might be expected that demonstrating productive knowledge would be more difficult than demonstrating receptive knowledge, in the current study, passive recognition of collocations was clustered closely with the most difficult active recall items of spelling and pronunciation based on hearing and reading the words respectively. This is of particular interest since the collocation items used for elicitation included three examples of collocates rather than just one. Moreover, unlike the current study, González-Fernández and Schmitt

(2020) found that recognition of collocates was the second most accurate item type of the eight aspects of word knowledge examined in their data. Though González-Fernández and Schmitt (2020) observed learner performance with recognition of collocates to be more accurate than performance with derivations, in the current study, learner performance with derivations was more accurate than with collocations.

Of course, a wide array of possibilities could account for these inconsistencies including different students learning in different contexts as well as the precise nature and differences of the instruments and elicitation processes. Since not all aspects of word knowledge can be studied at one time, our position is that many more studies need to be undertaken across as many components of word knowledge as possible. Then findings need to be aggregated to provide a general picture of the entire landscape. Though many scholars have aptly called for consistency in the ways in which vocabulary-based data are elicited to ensure comparability across studies, it is also important to note that there are many different types of data elicitation for a single aspect of word knowledge—each of which may be equally warranted for study. Thus, perhaps some focus needs to shift from simply labeling an aspect of word knowledge by the overarching terms such as definition, derivation, collocation, and so forth to a careful description of the specific elicitation contexts that includes the nature of the stimuli and the task. We may find that there may be many different ways to test particular aspects of word knowledge, each of which may occupy a different position in an implicational scale.

Teaching and Learning

The findings of this study coupled with previous research suggest a number of implications for L2 vocabulary development. First, it is imperative that practitioners and students understand the importance of vocabulary acquisition to L2 development and the unique challenges associated with L2 vocabulary acquisition (Barclay & Shmitt, 2019). Nation (1993) appropriately described the need for L2 learners to experience a flood of new vocabulary, particularly at lower proficiency levels. Moreover, practitioners and learners must understand which English vocabulary will be most important for their specific context. All who are learning English will benefit immensely from mastering the most frequent one thousand word families, which should provide more than 80% coverage of common texts (Nation, 2006).

While continuing to work toward mastery of the next few thousand most frequent word families, all learners are likely to benefit from learning academic vocabulary that is foundational to all disciplines such as found in the Academic Vocabulary List (Gardner & Davies, 2014). The organization

of this list is based on lemmas and includes part of speech, reducing many challenges associated with polysemy. At higher proficiencies, if learners have begun studying within specific disciplines, it may also be helpful for them to begin learning vocabulary from specialized lists of technical terms in fields such as business (Konstantakis, 2007), chemistry (Valipouri & Nassaji, 2013), engineering (Gustafsson & Malstrom, 2013), medicine (Wang, Liang, & Ge, 2008), and so on.

Since no single endeavor will provide all of the vocabulary development L2 learners need, Grabe (2009) has suggested that vocabulary learning must be advanced from multiple approaches simultaneously such as providing direct instruction to raise student awareness, helping students to apply effective vocabulary-learning strategies including using vocabulary notebooks or flashcards for ongoing review, learning new words through extensive reading, and ensuring students experience multiple encounters and ongoing recycling of new words. Once a robust effort toward vocabulary development is underway, findings from this and other studies suggest that students may benefit from learning experiences that initially emphasize receptive vocabulary knowledge and then move toward production such as pronunciation and spelling (also see Schmitt, 2019).

Vocabulary development is incremental over time (Barclay & Schmitt, 2019) and eventually learners should develop a deep understanding that includes knowledge of orthography, morphology, pronunciation, meanings, inflections, derivations, collocations, register, and so on. Nevertheless, in the short-term, particularly at lower proficiencies, effort should be made to minimize cognitive load on the learner while seeking to optimize vocabulary acquisition. Though perhaps counterintuitive for some, initially, this might take the form of learning more words (breadth) with fewer word-knowledge components rather than fewer words with more word-knowledge components (depth). This also might take the form of using L1 definitions, particularly at lower proficiency levels, to minimize cognitive load and expedite the speed and efficacy of vocabulary learning (e.g., Grace, 1998; Laufer & Shmueli, 1997; Nations, 1982). Many other efforts made by practitioners may support the vocabulary development of their students such as nurturing student motivation for vocabulary study and helping students to implement the most effective vocabulary review regimens (e.g., Barclay & Schmitt, 2019).

Limitations and Future Research

As with all research, limitations should be considered in the interpretation of these findings and in preparation for future research. First, though data were elicited from a substantial number of learners (110), this was about half the number of participants planned for this study. It is possible that a larger sample

of learners might have revealed greater differentiation of the relative difficulty of the aspects of word knowledge in the ANOVA and implicational scaling. Similarly, though the rationale for limiting word selection to the 2–3K range was to ensure that the different test forms functioned as similarly as possible, this range may have been too narrow for the smaller number of participants and may have adversely impacted the results. If large numbers of participants are not available, extending the frequency range of vocabulary studied might better reveal accuracy differences across word-knowledge components.

Conclusion

Building on the previous work of other scholars, this study used an innovative approach to creating an instrument designed to identify differential performance levels of ESL learners on eights aspects of word knowledge. Results showed performance levels varied across word-knowledge components and that implicational scaling was possible with some but not all aspects of word knowledge examined simultaneously. This study contributes to our understanding of important characteristics of vocabulary acquisition when examined in aggregate with other studies. Nevertheless, more research is needed to help clarify inconsistencies among studies. We believe that rather than limiting future research to traditional views of word-knowledge components, researchers should pursue the many different stimuli and tasks that could target a single aspect of word knowledge, thus greatly expanding our developing understanding leading toward a more complete theory of L2 vocabulary acquisition.

References

American Council on the Teaching of Foreign Languages (2012). ACTFL Proficiency Guidelines 2012. https://www.actfl.org/resources/actfl-proficiency-guidelines-2012

Atai, M. R., & Dabbagh, A. (2010). Exploring the role of vocabulary depth and semantic set in EFL learners' vocabulary use in writing. *Teaching English Language, 4*(2), 27–49. https://doi.org/10.22132/tel.2010.66106

Barclay, S., & Schmitt, N. (2019). Current perspectives on vocabulary teaching and learning. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 799–819). Springer International.

Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and*

language testing research (pp. 32–70). Cambridge University Press. https://doi.org/10.1017/CBO9781139524711.004

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693–713. https://doi.org/10.1093/applin/amq031

Davies, M. (2008). *Corpus of Contemporary American English (COCA).* Retrieved from https://www.english-corpora.org/coca/

Davies, M., & Gardner, D. (2010). *A frequency dictionary of American English: Word sketches, collocations, and thematic lists*. Routledge.

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning, 24*(1), 37–53. https://doi.org/10.1111/j.1467-1770.1974.tb00234.x

Ellis, R. (1988). The effects of linguistic environment on the second language acquisition of grammatical rules. *Applied Linguistics, 9*(3), 257–274. https://doi.org/10.1093/applin/9.3.257

Farvardin, M. T., & Koosha, M. (2011). The role of vocabulary knowledge in Iranian EFL students' reading comprehension performance: breadth or depth? *Theory and Practice in Language Studies, 1*(11), 1575–1580. https://doi.org/10.4304/tpls.1.11.1575-1580

Farvardin, M. T., & Valipouri, L. (2017). Probing the relationship between vocabulary knowledge and listening comprehension of Iranian lower-intermediate EFL learners. *International Journal of Applied Linguistics & English Literature, 6*(5), 273–278. http://doi.org/10.7575/aiac.ijalel.v.6n.5p.273

Frost, R. (2005). Orthographic systems and skilled word recognition. In M. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 272–95). Blackwell.

Gardner, D. & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305–327.

González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics, 41*(4), 481–505.

Grace, C. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning level CALL software. *The Modem Language Journal, 82(*4), 533–544.

Gustafsson, M., & Malmström, H. (2013). Master level writing in engineering and productive vocabulary: What does measuring academic vocabulary levels tell us? In N.-L. Johannesson, G. Melchers, & B. Björkman (Eds.), *Of butterflies and birds, of dialects and genres* (pp. 123–139). Stockholm Studies in English.

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*, 139–150.

Hakansson, G. (2013). Implicational Scaling. In P. Robinson (Ed), *Routledge encyclopedia of second language acquisition* (pp. 293–294). Routledge.

Hatch, E., & Lazaraton, A. (1991). *Design and statistics for applied linguistics: The research manuals*. Newbury House.

Hu, M. & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 23*(2), 403–430.

Kieffer, M. J., & Lesaux, N. K. (2012). Direct and indirect roles of morphological awareness in the English reading comprehension of native English, Spanish, Filipino, and Vietnamese speakers. *Language Learning, 62*, 1170–1204.

Kilic, M. (2019). Vocabulary knowledge as a predictor of performance in writing and speaking: A case of Turkish EFL learners. *PASAA: Journal of Language Teaching and Learning in Thailand, 57*, 133–164. https://files.eric.ed.gov/fulltext/EJ1224421.pdf

Koizumi, R. (2005). *Relationships between productive vocabulary knowledge and speaking performance of Japanese learners of English at the novice level*. [Doctoral Dissertation, University of Tsukuba]. University of Tsukuba Repository. https://tsukuba.repo.nii.ac.jp/?action=repository_action_common_download&item_id=20705&item_no=1&attribute_id=17&file_no=2

Konstantakis, N. (2007). Creating a business word list for teaching English. *Estudios de lingüstica inglesa aplicada, 7*, 79–102.

Laufer, B. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA review, 6*(1), 10–20.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). Macmillan. https://doi.org/10.1007/978-1-349-12396-4_12

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal, 28*(1), 89–107.

Laufer, B., & Goldstein (2004). Testing vocabulary knowledge: Size, strength and computer adaptiveness. *Language Learning*, *54*(3), 399–436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics, 36*, 611–634.

Meara, P. (1983). Vocabulary in a second language, vol. 1. *Specialised bibliography 3*. CILT.

Mehrpour, M., Razmjoo, S. A., & Kian, P. (2011). The relationship between depth and breadth of vocabulary knowledge and reading comprehension among Iranian EFL learners. *Journal of English Language Teaching and Learning, 53*(22), 97–127.

Milton J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Multilingual Matters. https://doi.org/10.21832/9781847692900-007

Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review, 63*, 127–147.

Nation, I. S. P. (1993). Measuring readiness for simplified material: A test of the first 1,000 words of English. In M. Tickoo (Ed.), *Simplification: Theory and application* (pp. 193–202). RELC.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I.S.P. (1982). Beginning to learn foreign vocabulary: A review of the research. *RELC Journal, 13*(1), 14–36.

Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22*, 31–55.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*(1), 1–28. https://files.eric.ed.gov/fulltext/EJ759833.pdf

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, *56*(2), 282–307. https://doi.org/10.3138/cmlr.56.2.282

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513–536. https://doi.org/10.1111/1467-9922.00193

Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21*(1), 28–52. https://doi.org/10.1191/0265532204lt273oa

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 41–60). Lawrence Erlbaum.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10*(1), 77–89. https://doi.org/10.2307/3585941

Rickford, J. R. (2002). Implicational scales. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (Eds), *The handbook of language variation and change*. Blackwell.

Schmitt, N. (1995). The word on words: An interview with Paul Nation. *Language Teacher, 19*(2), 5–7. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/1995-Interview-Norbert.pdf

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning, 48*(2), 281–317. https://doi.org/10.1111/1467-9922.00042

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge University Press. https://doi.org/10.1017/9781108569057

Schmitt, N. (2014). Size and depth of vocabulary knowledge: what the research shows. *Language Learning, 64*(4), 913–951. https://doi.org/10.1111/lang.12077

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*(2), 261–274.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition, 19*(1), 17–36. https://doi.org/10.1017/S0272263197001022

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

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*(2), 139–152. https://doi.org/10.1080/09571730802389975

Teng, F. (2014). Assessing the depth and breadth of vocabulary knowledge with listening comprehension. *PASAA: A Journal of Language Teaching and Learning, 48*(2), 29–56. https://files.eric.ed.gov/fulltext/EJ1077893.pdf

Thornbury, S. (2002). *How to teach vocabulary*. Pearson Education Limited.

Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes, 12*(4), 248–263.

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27*, 442–458.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*(01), 33–52. https://doi.org/10.1017/S0272263105050023

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Wesche, M., & Paribakht, T.S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *The Canadian Modern Language Review, 53*(1), 13–40. https://doi.org/10.3138/cmlr.53.1.13

Zhang, D., & Koda, K. (2017). Assessing L2 vocabulary depth with word associates format tests: Issues, findings, and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education, 2*(1), 1–30. https://doi.org/10.1186/s40862-017-0024-0

Zimmerman, K. J. (2004). The role of vocabulary size in assessing second language proficiency. [Master's thesis, Brigham Young University]. BYU Scholars Archive. http://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1577&context=etd

K. James Hartshorn, Aylin Surer

Zum Verständnis des Erwerbs von acht Aspekten der Vokabelkenntnisse

Zusammenfassung

Im Rahmen der vorliegenden Studie wurden auf Genauigkeit bezogene Daten von 110 ESL-Lernern erhoben – von der höheren Grundstufe bis zur niedrigen Oberstufe – mit der Absicht, einen Beitrag zu einer umfassenderen Theorie des Wortschatzerwerbs in der Zweitsprache zu leisten. Sie beziehen sich auf insgesamt 64 Vokabeln, die stichprobenartig aus der 2k-3k-Liste von COCA ausgewählt worden sind (32 Verben, 24 Substantive, 8 Adjektive) und acht Aspekte der Vokabelkenntnisse abdecken. Dazu gehören: die Rechtschreibung auf Grundlage der gehörten gesprochenen Form, die Wahl der Kollokationen auf Grundlage der geschriebenen Form, die Aussprache auf Grundlage der geschriebenen Form, die Wahl der Flexionsformen auf Grundlage des geschriebenen Kontextes, die Wahl der Definition auf Grundlage der gehörten gesprochenen Form, die Wahl der schriftlichen Definition auf Grundlage der geschriebenen Form, die Wahl entsprechender Ableitungen auf Grundlage der geschriebenen Form und die Wahl der geschriebenen Form auf Grundlage der schriftlichen Definition. Die ANOVA-Ergebnisse zeigen, dass das Genauigkeitsniveau bei verschiedenen Aspekten der Vokabelkenntnisse variiert sowie dass bei einigen, jedoch nicht bei allen simultan untersuchten Aspekten der Vokabelkenntnisse eine implizierende Skalierung möglich ist. In Zusammenhang mit anderen aktuellen und künftigen Studien bietet dies wichtige Schlussfolgerungen für die Entwicklung der Theorie des L2-Wortschatzerwerbs.

Schlüsselwörter: Wortschatzerwerb in der Zweitsprache, Aspekte des Wortwissens, implizierende Skalierung

Appendix A

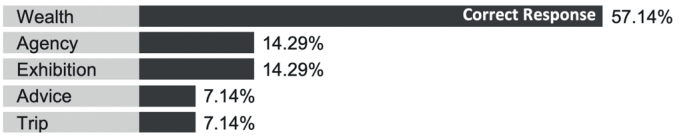
Words used in Data Elicitation Instrument (with frequency ranking)

1. accuse (2004)
2. mix (2091)
3. athlete (2169)
4. recover (2298)
5. philosophy (2345)
6. evaluate (2357)
7. wise (3046)
8. republic (2506)
9. question (2034)
10. approve (2098)
11. instrument (2112)
12. acquire (2331)
13. wealth (2351)
14. graduate (2407)
15. smooth (2903)
16. occasion (2530)
17. estimate (2042)
18. inspire (2118)
19. experiment (2011)
20. attract (2200)
21. academy (2474)
22. emphasize (2415)
23. rough (2847)
24. finance (2864)
25. invest (2048)
26. separate (2119)
27. revolution (2176)
28. divide (2239)
29. scholar (2493)
30. accomplish (2423)
31. straight (2434)
32. fiction (2607)
33. expose (2054)
34. reject (2128)
35. emotion (2178)
36. disagree (2261)
37. prince (2502)
38. adjust (2464)
39. brief (2463)
40. drama (2679)
41. convince (2056)
42. account (2147)
43. expense (2240)
44. compete (2291)
45. exception (2387)
46. assist (2467)
47. sharp (2408)
48. symbol (2780)
49. guide (2064)
50. assess (2157)
51. therapy (2303)
52. employ (2173)
53. passion (2388)
54. permit (2470)
55. pure (2391)
56. origin (2575)
57. pray (2070)
58. suspect (2165)
59. column (2315)
60. oppose (2192)
61. mystery (2398)
62. transform (2489)
63. rare (2015)
64. champion (2865)

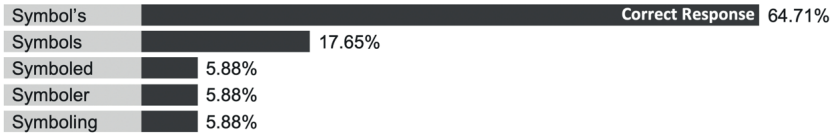
Appendix B

Additional Examples of Student Reponses to Multiple-Choice Items

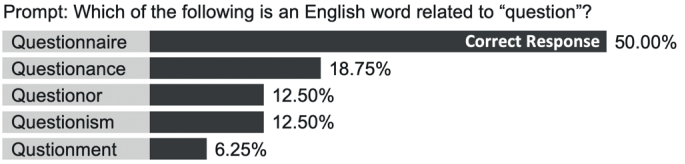
Select written form based on written definition
Prompt: Choose the word that matches the definition "a large amount of money or things someone has."



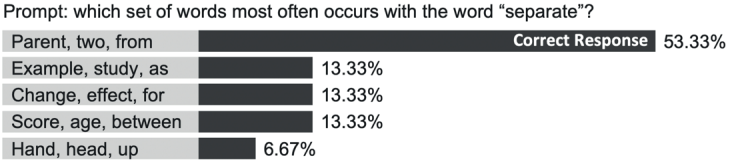
Select inflection based on written context
Use the right form of the word "symbol" in the following sentence. The _____ meaning changes between cultures.



Select derivation based on written form
Prompt: Which of the following is an English word related to "question"?



Select collocations based on written form
Prompt: which set of words most often occurs with the word "separate"?



Select definition based on hearing spoken form
Prompt: Click on the play button to listen to the recording: Choose the best definition of the word you hear.