Measuring Depth of Vocabulary Knowledge

Akifumi Yanagisawa and Stuart Webb

Introduction

Vocabulary research tends to focus on knowledge of form-meaning connection. Studies investigating the effects of teaching interventions on vocabulary learning often test whether participants can provide first language (L1) translations or synonyms for second language (L2) target words, or whether they can select corresponding L1 translations or synonyms for target words from among several options. Studies measuring the vocabulary size or breadth of learners’ knowledge also exclusively focus on whether test takers know the meanings of words. Measuring whether L2 learners know the meanings of L2 words is important without doubt; it may provide a better indication of whether learners can potentially understand words, compared to measuring other aspects of vocabulary knowledge such as spelling, pronunciation, and grammatical functions. However, the results of tests that only measure knowledge of the form-meaning connections of words may be misleading because vocabulary knowledge involves knowing many different things (e.g., derivations, collocations, associations) apart from form and meaning. Hence, how well words are known should also be considered.

Depth of vocabulary knowledge is typically defined as how well a word is known. Thus, measuring depth of vocabulary knowledge should tell us how well test takers know words. The importance of knowing vocabulary deeply is based mainly on two ideas: (1) L2 learners have to know different aspects of word knowledge in order to fulfill communicative tasks (i.e., reading, listening, writing, and speaking), and (2) advanced learners who can use vocabulary proficiently can demonstrate different aspects of knowledge of words. For example, knowledge of spelling is required for writing, knowledge of pronunciation is necessary for speaking, and knowledge of collocation is needed for both of these skills. Moreover, to use words appropriately, learners also have to know the contexts where specific words should and should not be used. Tests of vocabulary depth are thus of great value because they may signal the extent to which students have the knowledge to use vocabulary proficiently. Accordingly, measuring depth of knowledge is important for both educators and researchers. For educators, scores on depth measurements might indicate the extent to which L2 students have made progress in their vocabulary learning and the areas of vocabulary knowledge
(e.g., spoken form of vocabulary as opposed to written form) that require further attention. For researchers, scores on depth measurements may reveal how vocabulary knowledge develops and is associated with learning a language.

There are now many studies that have investigated depth of vocabulary knowledge. These studies vary greatly in their conceptualization and measurement of vocabulary depth. This chapter provides an overview of how depth has been defined and investigated. Moreover, the chapter will discuss important topics related to measuring depth of vocabulary knowledge and look at questions such as: To what extent can depth of vocabulary knowledge be measured? Which aspects of vocabulary knowledge should be measured? Which words to measure?

**Critical Issues and Topics**

**How Has Depth of Vocabulary Knowledge Been Defined?**

Depth of knowledge is often contrasted with breadth (or size) of knowledge, which refers to the number of words that are known. One oft-cited definition is provided by Anderson and Freebody (1981):

The first [vocabulary knowledge] may be called “breadth” of knowledge, by which we mean the number of words for which the person knows at least some of the significant aspects of meaning. . . . [There] is a second dimension of vocabulary knowledge, namely the quality or “depth” of understanding. We shall assume that, for most purposes, a person has a sufficiently deep understanding of a word if it conveys to him or her all of the distinctions that would be understood by an ordinary adult under normal circumstances. (pp. 92–93)

One important message from the conceptualization of depth by contrasting it with breadth is that in order to use the language competently, L2 learners need to not only know many words but also need to know each word deeply. It may be possible that some learners can provide the L1 translations for many L2 words by learning vocabulary with flash cards or word lists. In contrast, there may be other learners who do not know many words but have a deep understanding of words through engaging in different types of activities (e.g., extensive reading, writing activities, oral communication) that include a limited number of words (Schmitt, 2014).

Depth of vocabulary knowledge has been defined in many ways. For example, in Read (1993, p. 357), depth is described as “the quality of the learners’ vocabulary knowledge”. Wesche and Paribakht (1996, p. 13) defined depth “in terms of kinds of knowledge of specific words or in terms of degrees of such knowledge”. Henriksen (1999) conceptualized depth of knowledge as one of the three dimensions of lexical competence:

1. Partial to precise knowledge
2. Depth of knowledge
3. Receptive to productive use ability

The first dimension, partial to precise knowledge, refers to the strength of word knowledge. Henriksen’s description mainly focuses on the extent to which a word’s meaning or form is known. As to the second dimension, depth refers to different aspects of vocabulary
Measuring Depth of Vocabulary Knowledge

knowledge (e.g., various meanings of a word, associations among words, word class specification). She reports that to measure depth of knowledge, “researchers must ideally use a combination of test formats that tap [into] different aspects of knowledge” (p. 306). However, she discusses depth mainly in the context of network building, which is “the process of discovering the sense relations or intentional links between words – that is, fitting the words together in semantic networks” (p. 308). The third dimension, receptive to productive use ability, refers to the ability to understand and use a word. Henriksen suggests using various types of tests to measure each dimension of vocabulary knowledge.

Read (2004) noted that the term depth of knowledge has been used by many researchers in various ways. Read classified depth of knowledge into the following three components:

1. Precision of meaning
2. Comprehensive word knowledge
3. Network knowledge

The first and second categories, precision of meaning and comprehensive word knowledge, resemble the first and second dimensions of Henriksen’s (1999) conceptualization of lexical competence. Precision of meaning refers to how well the meaning of a word is known. Comprehensive word knowledge refers to multiple aspects of word knowledge. Read listed aspects of vocabulary knowledge as semantic, orthographical, morphological, syntactic, collocational, and pragmatic characteristics. The last category, network knowledge, refers to the degree to which words are organized or linked with each other in one’s mental lexicon.

In attempting to review earlier studies that focused on depth of knowledge in order to examine the relationship between depth and size (or breadth) of knowledge, Schmitt (2014) grouped depth of knowledge into seven conceptualizations: (1) receptive vs. productive mastery, (2) knowledge of multiple word knowledge components, (3) knowledge of polysemous meaning senses, (4) knowledge of derivative forms (word family members), (5) knowledge of collocation, (6) the ability to use lexical items fluently, and (7) the degree and kind of lexical organization.

As described earlier, depth of knowledge has been defined and conceptualized in many ways. Studies have adopted various conceptualizations of depth of knowledge, and accordingly, different conceptualizations require different approaches to measure depth of knowledge. The next section will look at how depth has been measured while referring to the corresponding conceptualizations of depth of knowledge described above.

How Has Depth of Vocabulary Knowledge Been Measured?

Ways to measure depth of knowledge can be categorized into three approaches: a developmental approach, a lexical network approach, and a components approach.

Developmental Approach

This approach operationalizes depth of vocabulary knowledge as developmental degree of word knowledge and considers that word knowledge develops from no knowledge to fully developed knowledge. Studies adopting the developmental approach use scales that indicate the developmental stage of vocabulary knowledge (Paribakht & Wesche, 1993; Schmitt & Zimmerman, 2002; Wesche & Paribakht, 1996). The most established and widely used test
adopting this approach is the Vocabulary Knowledge Scale (VKS: Paribakht & Wesche, 1993; Wesche & Paribakht, 1996). The VKS is designed to capture the developmental stage of word knowledge ranging “from complete unfamiliarity, through recognition of the word and some idea of its meaning, to the ability to use the word with grammatical and semantic accuracy in a sentence” (Wesche & Paribakht, 1996, p. 29). VKS’s five stages of word knowledge (Wesche and Paribakht call them categories) to measure depth are as follows:

I I don’t remember having seen this word before.
II I have seen this word before, but I don’t know what it means.
III I have seen this word before, and I think it means ________________. (synonym or translation)
IV I know this word. It means _______________. (synonym or translation)
V I can use this word in the sentence: ________________. (If you do this section, please also do Section IV.)

Test takers are asked to select the category that best represents how well one knows each word, and if possible, demonstrate their knowledge of the word. Points are provided based on their performance in the categories; when test takers select category I, one point is awarded, and when they use a word in a sentence with grammatical accuracy and semantic appropriateness, five points are awarded.

Although the VKS has been widely used in the field of vocabulary research, its validity has been frequently questioned (e.g., Bruton, 2009; Nation & Webb, 2011; Read, 2000; Schmitt, 2010; Webb, 2012; see also Kremmel, this volume). Among many issues that have been pointed out, the main critiques relate to two assumptions. First, the VKS assumes that the degree of difficulty between each category of the scale is equally spaced. However, this assumption is not based on any research findings, and different degrees of difficulty may be involved in advancing from one category to the next. For example, development from category I (i.e., I don’t remember having seen this word before) to category II (i.e., I have seen this word before, but I don’t know what it means) could be more easily achieved compared to the development from category II to category III (i.e., I have seen this word before, and I think it means ________). Usually studies use the combined VKS scores of tested words for each test taker; however, because the degrees of difficulty across categories are probably not equally spaced, the combined test scores cannot be appropriately interpreted. Second, the VKS also assumes that vocabulary knowledge develops linearly from category I (i.e., no knowledge of word form) to category V (i.e., words are used with grammatical accuracy and semantical appropriateness). Although this assumption seems reasonable, one can use a word with grammatical accuracy without clearly understanding its meaning (McNeill, 1996). Hence, this linear developmental assumption may not always be accurate. Despite these problematic assumptions, studies usually combine the VKS test scores across different categories, which obscures what type of vocabulary knowledge the test scores represent.

Lexical Network Approach

This approach conceptualizes depth of vocabulary knowledge as a lexical network in L2 learners’ mental lexicon, and research adopting this approach aims to inquire into the extent to which words are connected to each other (Henriksen, 1999; Meara, 1996, 2009; Meara & Wolter, 2004). This conceptualization corresponds to Read’s (2004) third definition of depth of knowledge, network knowledge. L2 learners’ lexical networks have been investigated
Measuring Depth of Vocabulary Knowledge

mainly by free word-association tasks (e.g., Fitzpatrick, 2013 for a review) as well as other passive association recognition tasks (e.g., Meara, 2009; Meara & Wolter, 2004; Wilks & Meara, 2002). Free word-association tasks ask examinees to provide the first word that comes to mind when hearing or seeing each cue word. It is believed that “[a] spontaneous response will reflect the strongest, or most salient, or most automatic link from that cue word” (Fitzpatrick, 2013, p. 6193). Building on word-association research, Read (1993, 1998) made the Word Associates Format (WAF) that requires learners to identify words that are associated with the target words instead of producing them. The WAF is arguably the most frequently used test format for measuring depth of vocabulary knowledge. Researchers have used the WAF to measure depth of knowledge and investigate the relationships between depth and breadth of knowledge, and/or between depth and language proficiency (e.g., reading, listening) (e.g., Qian, 1999, 2002; Qian & Schedl, 2004; Stæhr, 2009; Zhang & Yang, 2016; see also Qian and Lin, this volume). The following is an example of the WAF described in Read (1998).

sudden

<table>
<thead>
<tr>
<th>beautiful</th>
<th>quick</th>
<th>surprising</th>
<th>thirsty</th>
<th>change</th>
<th>doctor</th>
<th>noise</th>
<th>school</th>
</tr>
</thead>
</table>

The WAF involves learners selecting four words out of eight choices that have either a paradigmatic (have a related meaning) or syntagmatic (appear together in context, i.e., collocate) relationship with the cue word. In this example, quick and surprising are the paradigmatic associations for sudden, while change and noise are the syntagmatic associations for sudden.

The advantage of the WAF is that it measures multiple aspects of vocabulary knowledge, synonymy, polysemy, and collocations. However, the WAF also has several limitations. First, test takers can sometimes select appropriate associates without knowing a target word (Read, 1993; Schmitt, Ng, & Garras, 2011). Second, researchers’ use of the WAF varies from study to study in terms of the selection of cue words, association relationships, and test formats; therefore, research findings are not easily interpreted nor comparable across studies (see Zhang & Koda, 2017 for a recent review). Furthermore, studies administering the WAF tend to score the different associations as one variable, despite the WAF measuring different types of associations (i.e., paradigmatic and syntagmatic). This may limit the extent to which depth is revealed (Webb, 2012; Zhang & Koda, 2017). Lastly, because the WAF includes a focus on form-meaning connection, it might provide a rather narrow measurement of depth of vocabulary knowledge (Webb, 2012). For example, other dimensions of vocabulary knowledge, such as spelling, pronunciation, and grammatical functions, are also important for communicative tasks and would also be useful to measure.

Components Approach

This approach, also called the dimensions approach, conceptualizes depth of vocabulary knowledge by describing separate elements involved in knowing a word. This corresponds to Henriksen’s (1999) definition of depth of knowledge and Read’s (2004) second component of depth, comprehensive word knowledge. Among many others (e.g., Cronbach, 1942; Richards, 1976), Nation’s (2013) description of vocabulary knowledge is the most comprehensive and widely cited classification of the different dimensions (Table 24.1).

Nation (2013) classified vocabulary knowledge into three components: form, meaning, and use. Each component is broken down into three aspects of knowledge. Each aspect
is further broken down into receptive and productive knowledge. Receptive (or passive) knowledge is the knowledge that is required to comprehend words when encountered during reading or listening. Productive (or active) knowledge is what is required to use words for speaking or writing. Studies adopting this components approach have focused upon measuring one or several of these aspects of vocabulary knowledge.

In the components approach, it is very important to clearly distinguish depth of knowledge and strength of knowledge (Nation & Webb, 2011; Webb, 2012). Webb (2012, p. 3) defines strength of knowledge as “how well a single aspect is known”. For example, the written form of a word (i.e., spelling) can be known to different degrees, from no knowledge to degrees of partial knowledge, which may include being able to identify the word when seeing it or being able to write the word with an inaccurate but identifiable spelling, to full knowledge, where one can quickly produce the complete and exact spelling of a word. In Laufer and her colleagues’ studies (Laufer, Elder, Hill, & Congdon, 2004; Laufer & Goldstein, 2004), the term strength of knowledge is used to compare the different degrees of knowledge of form-meaning connections: receptive (or passive) recognition and recall, and productive (or active) recognition and recall. Strength of knowledge may also refer to fluency of word knowledge, i.e., how quickly one aspect of word knowledge can be accessed (see also Godfroid, this volume, for how different aspects of word knowledge can be measured with tacit-implicit, automatized explicit, or procedural methods). Strength of knowledge is thus one important indicator of the quality of word knowledge. However, strength and depth of knowledge should not be considered as equivalents. Because these two terms refer to two different qualities of word knowledge, treating them as one construct complicates the term depth and may reduce the generalizability of findings across studies. For example, Vermeer (2001) operationalized depth as how well the meanings of words are known when investigating the relationship between depth and breadth of knowledge. The results showed a strong

### Table 24.1 Description of word knowledge

<table>
<thead>
<tr>
<th>Form</th>
<th>Spoken</th>
<th>R</th>
<th>What does the word sound like?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>How is the word pronounced?</td>
</tr>
<tr>
<td>Written</td>
<td>R</td>
<td>What does the word look like?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>How is the word written and spelled?</td>
</tr>
<tr>
<td>Word parts</td>
<td>R</td>
<td>What parts are recognizable in this word?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>What word parts are needed to express the meaning?</td>
</tr>
<tr>
<td>Meaning</td>
<td>Form and meaning</td>
<td>R</td>
<td>What meaning does this word form signal?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>What word form can be used to express this meaning?</td>
</tr>
<tr>
<td>Concept and referents</td>
<td>R</td>
<td>What is included in the concept?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>What items can the concept refer to?</td>
</tr>
<tr>
<td>Associations</td>
<td>R</td>
<td>What other words does this make us think of?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>What other words could we use instead of this one?</td>
</tr>
<tr>
<td>Use</td>
<td>Grammatical functions</td>
<td>R</td>
<td>In what patterns does the word occur?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>In what patterns must we use this word?</td>
</tr>
<tr>
<td>Collocations</td>
<td>R</td>
<td>What words or types of words occur with this one?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>What words or types of words must we use with this one?</td>
</tr>
<tr>
<td>Constraints on use</td>
<td>(register, frequency . . .)</td>
<td>R</td>
<td>Where, when, and how often would we expect to meet this word?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>Where, when, and how often can we use this word?</td>
</tr>
</tbody>
</table>

**Note:** R = receptive knowledge, P = productive knowledge.

**Source:** Adapted from Nation (2013, p. 49)
Measuring Depth of Vocabulary Knowledge

correlation between the two. This led Vermeer to claim that depth and breadth are the same construct. However, as Nation and Webb (2011) and Schmitt (2014) point out, looking only at one aspect (in this case form-meaning connection) may not be enough to reveal comprehensive depth of vocabulary knowledge. Hence, it would be more meaningful to distinguish between the quality of one aspect of word knowledge from the quality of multiple aspects of word knowledge by referring to the former as strength of knowledge and the latter as depth of knowledge.

Studies using the components approach can be further divided into two groups, studies measuring multiple aspects of knowledge and studies focusing on one single aspect of word knowledge (e.g., collocation, derivatives). The former studies conceptualize depth of knowledge as the degree to which different components are known (or learned) while measuring the same set of words. For example, Schmitt (1998) tracked three advanced English learners’ acquisition of 11 words over a year. He assessed the learner’s strength of knowledge of spelling, associations, grammatical information, and polysemy for the words. The results showed that multiple aspects of vocabulary knowledge developed gradually. While knowledge of spelling was high from the beginning, learners rarely knew all of a tested word’s multiple meanings or derivative forms. In a series of studies, Webb investigated the degree to which different learning conditions contributed to vocabulary depth by measuring five different components of knowledge: orthography, syntax, association, grammatical functions, and form-meaning connection (Webb, 2005, 2007a, 2007b, 2009a, 2009b). Each component was measured using receptive and productive tests (recognition and recall formats, respectively). For example, Webb (2007b) demonstrated that different aspects of word knowledge can be learned incidentally through reading. Furthermore, frequency of encounters influenced the development of different aspects of word knowledge differently.

One advantage of measuring multiple aspects of vocabulary knowledge is that test scores may be more meaningful. When using one test that measures multiple aspects of vocabulary knowledge together, it may not be clear what type of vocabulary knowledge the test score represents. Another advantage is that scores on different tests can be directly compared with each other because every test measures the same set of words. This allows researchers to examine how comprehensively different components of word knowledge are known and the hierarchical order of acquisition of various components of a word. However, measuring different components of knowledge does have some limitations. For example, measuring multiple aspects of vocabulary knowledge tends to take a long time (Nation & Webb, 2011; Schmitt, 1998), which increases the burden on both test takers and researchers. Accordingly, only small numbers of words are focused on in each individual study. Test results based on the measurement of a small number of words may not adequately describe L2 learners’ comprehensive vocabulary knowledge (Meara & Wolter, 2004; Read, 2000; Schmitt, 2010).

The second group of studies that might be classified under the components approach have focused on providing a more accurate measurement of one aspect of knowledge rather than measuring several aspects. Quite frequently measurements focusing on vocabulary knowledge, such as collocations, word parts, and grammatical functions, have been referred to as depth tests even when measurement focused only on one aspect of word knowledge (e.g., Gyllstad, 2013; Henriksen, 1999; Milton, 2009; Schmitt, 2000, 2014).

For example, several tests have been developed to measure knowledge of collocations (Gyllstad, 2009; Nguyen & Webb, 2017; see also Gyllstad, this volume), and word parts (or derivatives) (Sasao & Webb, 2017; Schmitt & Zimmerman, 2002; see also Sasao, this
volume). The following is an example of the collocation test format administered by Nguyen and Webb (2017).

1 money  a. check  b. drop  c. make  d. miss
2 figure  a. central  b. funny  c. happy  d. usual

In this example, test takers were asked to choose the most suitable verb (make) and adjective (central) that can be combined with noun prompt words. Nguyen and Webb sampled collocations from three frequency levels (1,000, 2000, and 3,000) when selecting target items.

Schmitt and Zimmerman’s (2002) Test of English Derivatives taps into knowledge of different derivative forms within major word classes (i.e., noun, verb, adjective, and adverb). The following is one example. The words in parentheses are the answers.

1 stimulate
Noun A massage is good ___________. (stimulation)
Verb Massages can ________________ tired muscles. (stimulate)
Adjective A massage has a ________________ effect. (stimulating)
Adverb He massaged _________________. (X)

Test takers are to write the correct form of the cue word in each sentence. If there is more than one possible answer, test takers are to write only one, and if there is no form, to write X.

Tests measuring one specific aspect of vocabulary knowledge are helpful to reveal an L2 learner’s strength of knowledge for a specific aspect. If a more common test of form-meaning connection, such as the Vocabulary Levels Test, is also administered then together they may provide some indication of depth. However, one limitation would be that because each test is developed separately, the test scores may not necessarily be comparative across different aspects. Focusing only on one aspect of word knowledge should also be considered a very narrow measurement of vocabulary depth. To overcome this limitation, Ishii and Schmitt (2009) proposed a principled approach to diagnosing vocabulary weakness by comparing student performance to the norm of their peers. In their study, Japanese EFL students’ vocabulary size and knowledge of multiple meanings, derivatives, and lexical choice between near-synonyms were measured. Each student’s scores on different vocabulary tests were presented in a radar chart, by which a student can easily understand the area of vocabulary knowledge that they should focus on. Ishii and Schmitt could successfully detect variations in different vocabulary measurements. It would also be beneficial for researchers to develop standardized tests, in which different aspects of word knowledge can be directly compared with each other.

Which Approach Should We Use to Measure Depth of Knowledge?

As described earlier, researchers have conceptualized depth in several ways and have used different approaches to measure the quality of L2 learners’ vocabulary knowledge. In order to disentangle the definition of depth of vocabulary knowledge, it is beneficial to discuss which approach works best.

The first approach is the developmental approach, which examines depth of word knowledge by looking at the developmental stage of word knowledge. However, as seen in many criticisms of the VKS, combining different aspects of word knowledge in a single test may...
reduce the validity of the test by making it unclear which type of knowledge the test scores reflect. One might argue that the developmental approach can be used by focusing on one aspect of word knowledge and tracking its development (e.g., from no knowledge to full knowledge of the spelling of a given word). Although this is a useful measurement, it may be better treated as a measure of a single component rather than depth, unless it is measured together with other aspects of knowledge. It is also quite different from what many researchers consider as depth of knowledge. Instead, it might be best to consider it a measure of strength of knowledge (Nation & Webb, 2011; Webb, 2012).

Another approach is the lexical network approach, which examines the L2 learners’ mental lexicon. Read’s (1993, 1998) WAF has been used to indicate one’s lexical network or knowledge of paradigmatic and syntagmatic connections among words. For example, synonyms, near-synonyms, and collocations are treated as an indication of one’s depth of vocabulary knowledge. However, two main issues were pointed out. First, because the WAF combines scores of different types of word connections, it is not clear which aspects of vocabulary knowledge the test scores represent. Second, the degree of depth of knowledge that the WAF taps into is quite narrow and may not be sufficiently comprehensive to claim that the test reflects one’s comprehensive lexical network or depth of knowledge. Although the WAF represents an important development in vocabulary testing, within the lexical network approach we might hope for the development of a test that taps into more aspects of vocabulary knowledge.

This leaves the components approach as the last plausible path. The advantage of the components approach is that each measurement has high construct validity (i.e., measuring the construct that a test is supposed to be addressing) because it focuses on each aspect of word knowledge separately. Because of this, test scores are more easily and clearly interpreted, compared to using one test that claims to be measuring depth of vocabulary knowledge. Additionally, each aspect can be investigated more deeply by using different levels of sensitivity to measure its strength of knowledge. It is also necessary to simultaneously measure different aspects of word knowledge in order to reveal a comprehensive picture of L2 learners’ quality of vocabulary knowledge. Recent advancements in the use of statistical analysis in the vocabulary research field can also broaden research on depth of knowledge. For example, Structural Equation Modeling (SEM) allows us to examine multiple aspects of vocabulary knowledge in an integrated manner (Kieffer & Lesaux, 2012; Koizumi & In’nami, 2013; Tseng & Schmitt, 2008; Zhang, 2012). Mixed-effects statistical models enable us to examine the influence of the characteristics of words (e.g., cognate status, pronounceability) and the characteristics of examinees (e.g., proficiency, L1 background, learning contexts) on test performance simultaneously (e.g., Cunnings, 2012). In addition to the use of traditional approaches (e.g., statistical inference to compare mean differences, multiple regression analysis), advanced statistical analyses may further help us investigate the complicated relationships between proficiency and different aspects of vocabulary knowledge and between size, depth, and lexical development.

To What Extent Can Depth of Vocabulary Knowledge Be Measured?
It may not be realistic to provide a comprehensive measure of vocabulary depth. The reason behind this is that even when using a large battery of tests or an interview, we may only measure a fraction of knowledge of a word. Imagine a case in which we would like to measure knowledge of the word make. We could ask students to provide an L1 translation for make. However, make has multiple meanings. For example, make dinner, make money, make a
reservation, and make one’s bed refer to different actions. One can also express different meanings by combining make with other words, such as I didn’t make it to the train last night, make up your mind, or I could not make out what she was saying. Furthermore, knowing the written form of make does not necessarily mean that one knows the spoken form, and vice versa. If we want to measure knowledge of collocation, there are a large number of useful words that co-occur with make, such as sure, sense, difference, money, decision(s), feel, clear, easier, changes, happen, living, and many more. However, it would be difficult to measure knowledge of more than a small number of collocations for a word. Similarly, if we want to measure knowledge of the derivations of make, then we would hope to reveal whether test takers know makes, making, makings, made, maker, remake, remade, unmake, and unmade. As this example of make illustrates, there is a great deal involved in knowing a word, and it may be very challenging to determine the extent to which everything is known about a word. Even in studies measuring multiple aspects of word knowledge, a fraction of a student’s knowledge may be measured. One way to deal with this difficulty is to focus on the strength of knowledge regarding specific aspects of depth of knowledge in relation to the purpose of the study.

Which Aspects of Vocabulary Knowledge Should Be Measured?

In research and pedagogy, it is important to consider which aspects of vocabulary knowledge to measure. The aspects of vocabulary knowledge that are measured should be selected in relation to the teaching and learning purpose. If we are interested in the degree to which words are learned in different activities, the aspects that the activities contribute to should be measured. For example, crossword puzzles, matching activities, and multiple-choice questions typically contribute most to learning form-meaning connection and written form, so perhaps these two aspects might be measured. Meaning-focused reading and listening activities contribute to most aspects of knowledge and so may merit measuring knowledge of aspects such as grammatical functions, word parts, collocations, and form-meaning connection. Perhaps of most interest in activities that involve the use of words in speech and writing would be measuring knowledge of collocations and associations because these aspects are of particular importance in production. However, it would also be meaningful to measure spoken and written word form when comparing vocabulary learning from reading, reading while listening, and listening activities. For example, reading while listening activities might better facilitate learning of spoken word forms (e.g., pronunciation) in addition to written word forms.

Which Words to Measure?

Which target words to measure should be based on the purpose of the study. When trying to draw a comprehensive picture of L2 learners’ vocabulary knowledge, it is probably helpful to measure knowledge of words at different frequency levels. Frequency is considered to be one of the most influential factors contributing to language acquisition (e.g., Ellis, 2002). More frequent words tend to be learned before less frequent words (e.g., Webb, Sasao, & Ballance, 2017). Additionally, learners may have deeper knowledge of words at higher frequency levels compared to those at lower frequency levels (Schmitt, 2014). For example, Nguyen and Webb (2017) found that Vietnamese EFL learners’ receptive knowledge of collocations tended to develop according to the frequency of the individual words in the collocations; collocation knowledge increased as the frequency of the individual words
comprising the collocations increased. However, to date, few attempts have been made to look at depth of knowledge of words at different frequency levels (Greidanus, Bogaards, van der Linden, Nienhuis, & de Wolf, 2004; Nguyen & Webb, 2017). Measuring words at different frequency levels would be beneficial because this may indicate (1) up to which frequency level learners have acquired a specific component of vocabulary knowledge, and (2) which frequency level learners should focus on to foster their depth of knowledge. Furthermore, longitudinally tracking multiple components of word knowledge at different frequency levels may show the developmental speed of each component of word knowledge, which in turn would help to reveal a more comprehensive picture of how vocabulary knowledge develops.

In studies examining how different aspects of word knowledge are developed through language learning (e.g., Pigada & Schmitt, 2006; Schmitt, 1998; Webb, 2005, 2007a, 2007b, 2009b), words that participants encounter during a learning condition are best suited to be tested. Target words can be the words that participants do not know, or words that are partially known; some components of word knowledge may be known while other components may not be known.

Future Directions

There are three key areas for further research on depth of vocabulary knowledge. The first is to examine how different components of depth relate to each other and to vocabulary size. Research demonstrates that different measures of depth of knowledge tend to correlate with each other (e.g., Ishii & Schmitt, 2009; Schmitt & Meara, 1997). Furthermore, studies tend to find a high correlation between tests measuring different components of depth and size tests that measure knowledge of the form-meaning link (e.g., Gyllstad, 2009; Qian, 1999; Tannenbaum, Torgesen, & Wagner, 2006; Vermeer, 2001). These findings may indicate that some aspects of vocabulary knowledge develop simultaneously while others develop at different rates. The high correlation between size and depth measurements indicates that there may be little value in measuring depth of knowledge (Milton, 2009; Schmitt, 2014). However, as mentioned in Schmitt’s (2014) comprehensive review of earlier studies measuring both size and depth, the correlation between size and depth varies from study to study. Hence, it is still not clear how different aspects of vocabulary knowledge relate to each other. The size of the correlation may be dependent upon variables such as test format, participant differences, and study design. It may be worth conducting individual studies or meta-analyses to investigate when correlations between depth and size increase and/or decrease. Similarly, looking at the relationships between different aspects of word knowledge may reveal how closely they are related. However, it is also important to note that if measurements highly correlate with each other, those measures may merely be different tests measuring the same construct. On the other hand, lower correlations between different aspects of knowledge may signal constructs that researchers should not combine.

The second area where research is needed is examining how depth of knowledge contributes to language teaching. Researchers have emphasized the pedagogical importance of depth of vocabulary knowledge (e.g., Nation & Webb, 2011; Schmitt, 2014; Webb, 2012) suggesting that language learners should not only focus on learning form-meaning links but also try to learn multiple aspects of knowledge of a word in order to use it in a communicative context. However, few studies have focused on how measures of depth of vocabulary knowledge benefit language teaching. Ishii and Schmitt (2009) pointed out that there is little practical advice on how depth test scores can be interpreted for pedagogical purposes. They proposed a principled approach to diagnosing vocabulary weakness by comparing student
Akifumi Yanagisawa and Stuart Webb

performance to the norms of their peers. In order to increase the value of such an approach, it is also important to investigate to what extent each aspect of knowledge relates to students’ actual use of words.

Two lines of research could conceivably deepen our understanding of the relationship between how well words are known and how well the words can be used. The first is to investigate how well scores on different depth of knowledge measurements explain students’ language proficiency (i.e., reading, writing, listening, and speaking). Research has demonstrated that vocabulary size largely predicts learners’ language abilities (e.g., Milton, 2010, see also Qian and Lin, this volume). Some researchers looked at the relationship between language ability and vocabulary knowledge by measuring both size and depth (Kieffer & Lesaux, 2012; Koizumi & In’nami, 2013; Qian, 1999, 2002; Qian & Schedl, 2004; Stæhr, 2009; Tannenbaum et al., 2006; Zhang & Yang, 2016). The findings show that depth and size tests tend to highly correlate with each other, and by and large depth and size tests provide similar predictive power of learners’ proficiency. For example, Qian (1999) looked at the relationship between reading comprehension and size and depth of vocabulary knowledge. Depth of vocabulary knowledge was measured with Read’s WAF. The results showed that depth added a further 11% to the prediction of reading comprehension over and above the prediction already provided by vocabulary size. Based on this, Qian concluded that not only size but also depth is an important aspect of vocabulary knowledge for predicting reading comprehension. On the other hand, Stæhr (2009) examined the relationships between vocabulary size, depth, and listening comprehension, and found that depth only added 2% to the prediction of listening comprehension on top of the prediction already made by size. The contribution of depth of knowledge may differ based on the component of depth measured and on the characteristics of the language ability measured. Zhang and Yang (2016) investigated learners of Chinese and found that vocabulary size was a more important predictor of comprehension of longer passages than depth of knowledge, while depth was a better predictor of comprehension of shorter passages. So far, a general agreement of whether depth (and depth components) uniquely contributes to the prediction of language ability over and above vocabulary size has not been reached, making further research warranted. Results from further research may indicate (1) which aspects of vocabulary knowledge are important to certain language abilities, and (2) which aspect of depth of knowledge teachers should measure to get an idea of how proficient students are in a specific language ability.

The second line of research on the pedagogical value of depth of knowledge would be to measure different aspects of vocabulary knowledge that are developed through completing different types of language activities. It is reasonable to assume that activities in which learners focus on form-meaning connection exclusively develop learners’ knowledge of this aspect. Interestingly, two of Webb’s (2007a, 2009b) studies investigating the effects of different types of learning activities provided counterintuitive results. Webb (2007a) investigated the effects of learning word pairs and learning from a single glossed sentence. The results indicated that the provision of context did not contribute to greater learning of different aspects of vocabulary knowledge. In Webb’s (2009b) study, EFL students learned L2 vocabulary using word lists. A battery of ten tests measuring several aspects of vocabulary knowledge revealed that students can learn not only form-meaning connection but also collocation, grammatical functions, and associations to some extent just through linking L1 translations to target words. Therefore, it would be useful to further investigate which types of activities foster different aspects of vocabulary knowledge.

The last research area that requires more attention is the development of standardized tests that tap into each aspect of vocabulary knowledge. There are already many tests developed
and recommended for measuring different aspects of vocabulary knowledge (e.g., Gyllstad’s (2009) COLLEX and COLLMATCH tap into knowledge of collocation; Sasao and Webb’s (2017) Word Part Levels Test measures knowledge of word parts). However, because existing measurements are developed independently, scores on the tests cannot be directly compared to each other. It would also be useful to develop a battery of tests designed to measure different aspects of vocabulary knowledge of words at different frequency bands (Milton & Hopkins, 2006; Nation & Webb, 2011; Nguyen & Webb, 2017). Nguyen and Webb (2017) measured collocational knowledge of words at the first three 1,000-word frequency levels, as well as single-word items at the same frequency levels. Milton and Hopkins (2006) measured students’ knowledge of written and spoken form-meaning connection by using two different test formats (i.e., X-Lex & AuralLex). Similar methods can be applied to develop a set of vocabulary tests that measure other aspects of vocabulary knowledge such as derivational forms, semantic associations, and polysemy. The results of this test battery may reveal the quality of students’ knowledge of words at different frequency levels. The test scores could also be directly compared to each other and could be interpreted much easier.

**Further Reading**


This study investigated how encountering words in 1, 3, 7, and 10 sentences contributed to different aspects of vocabulary knowledge. The article described how different components of depth of vocabulary knowledge can be measured and interpreted.


The chapter on measuring depth of vocabulary knowledge describes the steps that researchers should take when measuring depth of knowledge. It also provides a practical guide to measuring depth of knowledge.


The chapter on measuring vocabulary knowledge provides a review of previous studies focused on measuring depth of knowledge and discusses the advantages and issues related to different test formats.

**Related Topics**

Vocabulary assessment, vocabulary tests, quality of word knowledge, vocabulary size, lexical development, measuring vocabulary learning progress

**References**


Measuring Depth of Vocabulary Knowledge


Akifumi Yanagisawa and Stuart Webb


