Reading in a Foreign Language April 2021, Volume 33, No. 1
ISSN 1539-0578 pp. 126–140

The coverage comprehension model, its importance to pedagogy and research, and threats to the validity with which it is operationalized

Stuart McLean
Momoyama Gakuin University
Japan

Abstract

When learners can comprehend 98% or more of the tokens within a text, the lexical difficulty of the text is unlikely to inhibit reading comprehension (Schmitt et al., 2011). This phenomenon will be referred to as the Coverage Comprehension Model (CCM). The CCM is present in countless articles that describe the percentage of tokens necessary to comprehend reading materials (e.g., Nation, 2006). Further, numerous studies operationalize the CCM to provide evidence that participants were able to comprehend reading materials (e.g., Feng & Webb, 2020) by estimating (a) the lexical difficulty of a text and (b) the lexical mastery level of a learner. However, the validity with which the CCM is operationalized is limited by the following four assumptions; (a) 26 out of 30 words on a levels test is an appropriate threshold for mastery of a 1,000-word band; (b) the word counting unit used when estimating the lexical difficulty of a text and the lexical ability of a learner is appropriate for the target learners; (c) the item format used in levels tests can appropriately capture the type of vocabulary knowledge necessary when reading; and (d) the number of items on a vocabulary levels test accurately represents the difficulty of the 1,000-word band. This paper applies the findings of research to evaluate the validity of the first two assumptions, and concludes that the validity with which the CCM is operationalized in research is limited.

***Keywords*:** Text coverage model, lexical coverage, reading comprehension, vocabulary knowledge, vocabulary learning, validity

Introduction

This review of the application of the Comprehension Coverage Model (CCM) within reading research was conducted to improve research methodology, which can be achieved through the application of novel research methods. However, a simpler method for improving research robustness is to avoid limited research practices. Moreover, errors are best avoided by first highlighting and noticing them.

The text coverage model

If learners know 98%[1] or more of the tokens within a text, the lexical difficulty of the text should not inhibit reading comprehension (Hu & Nation, 2000; Laufer, 1989; Schmitt et al., 2011). Hereafter, this phenomenon is referred to as the Coverage Comprehension Model (CCM). The CCM phenomenon is operationalized in countless articles that describe the percentage of tokens (running words) necessary to comprehend reading materials (e.g., Gui et al., 2020; Nation, 2006, 2014; Feng & Webb, 2020) and listening materials (e.g., Nation, 2006; Feng & Webb, 2020). Further, numerous articles operationalize the CCM to provide evidence that the reading materials used were comprehensible to their participants (e.g., Feng & Webb, 2020; Huffman, 2014; Waring & Takaki, 2003). Because the CCM is the cornerstone of research and pedagogy concerning the importance of vocabulary to reading, the valid operationalization of the CCM by establishing the lexical difficulty of the text and the lexical mastery level of the learner is critical.

The application of the text coverage model

There are two challenges when matching learners with reading (or listening) materials of an appropriate level in English as a foreign language (EFL) or expanding circle education settings, where most learners share the same first language (L1) and the generally low proficiency learners have limited exposure to English. The first challenge is establishing the lexical mastery level of the student. A vocabulary levels test (*hereafter levels test*) is better for this purpose than vocabulary size tests (McLean & Kramer, 2015; Nation, 2016; Stoeckel et al., 2020). Levels tests are based on a corpus-derived frequency-based word list. It is not possible to test learners on each word in a 1,000-word band[2]. Instead, each band is represented by randomly selected words (20-40). For each word, a vocabulary item (question) is created. Usually, the target word form in the item is the word's base form (i.e., *use*) and not a derivational form (i.e., *usage*). Finally, mastery of a 1,000-word band is determined by correctly answering a stated percentage of the items in a band.

The second challenge is establishing the lexical load of a reading text (or editing a text so that it has a desired lexical load). When teachers write or edit the lexical load of reading (and listening) materials to a 1,000-word level or investigate the lexical load of the materials, they usually use a word profiler to indicate which words are outside the target 1,000-word range. For example, to write passages to the 2,000-word level, teachers must find or edit a text so that 98% of its tokens are from the first two 1,000-word bands.

Once these two challenges have been overcome, a learner can be matched with materials that are written or edited to the same level. For example, a learner who demonstrates mastery of the first two 1,000-word bands can be given a text in which 98% of the tokens are from those two bands.

[1] This article uses 98% while acknowledging that other figures have been suggested for different reading purposes.
[2] Hereafter only 1,000-word bands will be referred to, however, especially for lower-level learners and high-frequency words, 100- to 500-word band analysis is of value.

Threats to the application of the text coverage model

The inferences that (a) a text is of an estimated lexical load, (b) a learner has mastery of the stated 1,000-word bands, and (c) a text is lexically appropriate for a learner, are based on four major assumptions. First, the threshold for mastery of a 1,000-word band is knowledge of 26 out of 30 words on a levels test. Second, the word counting unit used for creating the frequency-based word list is appropriate for the target learners. Third, the item format used in levels tests appropriately captures the type of vocabulary knowledge needed for reading. Fourth, the number of items selected to represent a 1,000-word band accurately represents that band. This paper applies the findings of research to evaluate the validity of the first two assumptions. See McLean et al. (2020) and Zhang & Zhang (2020) for discussions of the third point, and Stoeckel et al. (2020) and Gyllstad et al. (2020) for discussions of the fourth point.

Paradigm shift: Justifying methods with research findings and not convention

Vocabulary testing is often based on past practice, and the testing purpose or construct validity is rarely considered (Schmitt et al., 2020: Stoeckel et al., 2020). This has been the case when interpreting levels tests scores and mastery thresholds in reading research (Xing & Fulcher, 2007).

Justifying research methodology with inappropriate citations. In reading research, mastery of a 1,000-word band is commonly set at 26/30 words, which is 86.6%. The 86.6% threshold figure has been used in conjunction with levels tests to argue that materials were lexically appropriate (Feng & Webb, 2020; Huffman, 2014). For example, Feng and Webb adopted the 86.6% threshold to investigate learners' mastery of 1,000-word bands, justifying this threshold only by reference to Schmitt et al. (2001). First, Feng and Webb argued that around 95% knowledge of tokens within a text is necessary for comprehension. Then, they calculated that the first three 1,000-word bands provided 95% coverage of the target treatment materials. Finally, they decided that the lexical coverage level of the transcript was appropriate[3] for the participants in each group based on their performance on a levels test. Thus, Feng and Webb state that learners need to be able to comprehend around 95% of tokens for comprehension, and establish which 1,000-word band provides around 95% coverage of the target material. However, Feng and Webb then establish which 1,000-word band learners have mastery of using the figure of not 95%, but only 86.6%. If Feng and Webb had used the mastery threshold of 29/30 (96.67%), a mastery threshold suggested by Webb et al. (2017), then only 94.7%, 46.1%, and 21.1% of the participants would have demonstrated mastery of the first, second and third 1,000-word bands, rather than the reported 100/%, 92.1%, and 60.5% (Y. Feng, personal communication, January 2, 2021).

Xing and Fulcher (2007) note that when discussing the interpretation of levels test scores, mastery thresholds are not supported by research. One reason why 86.6% became the standard mastery threshold is that the Vocabulary Levels Test (VLT) proposed by Schmitt et al. (2001)

[3] It is assumed that "appropriate" means appropriate for comprehension and incidental vocabulary learning as Feng and Webb's study looked at incidental vocabulary acquisition through reading, listening, or reading while listening. Feng and Webb state that "[i]ncidental learning occurs when unknown words are encountered repeatedly in meaning-focused input" (2020, p. 7). Webb and Nation (2017) state that "[a]s with listening, reading activities are classified as meaning-focused input when there is a focus on comprehension and a low density of unknown words (2% or less)" (2017, p. 78).

established 86.6% as the mastery threshold and this levels test became perhaps the most widely used measure of L2 lexical knowledge. Another reason is that the 86.6% figure became commonly used in research, even though the only rationale provided is often reference to Schmitt et al. (2001).

Applying research when deciding purpose-specific mastery thresholds. Schmitt et al. (2001, p. 67) state that "[l]ike Read, we carried out a Guttman scalability analysis (Hatch and Lazaraton, 1991), using a criterion of mastery of 26 out of the 30 possible per level. (This figure was chosen to be as close as possible to Read's criterion of 16 out of 18.)". Read (1988, p. 17) looked at the scalability of scores from the 1,000-word bands of Nation's (1983) levels tests and states that "[a] score of 16 was taken as the criterion for mastery of the vocabulary at a particular level." Read "set the cut score at 16/18 based on [his] reading on criterion-referenced testing at the time, which indicated that a score equivalent to 90% was widely accepted as the criterion for mastery, so 16/18 represented 90% for a VLT level" (J. Read, personal communication, January 28, 2021). However, this criterion-based research was not related to the lexical knowledge necessary for reading. Thus, "contemporary vocab researchers need to revisit the mastery cut-off in light of recent developments in the field, their research aims and the targeted purposes for reading, rather than just quoting Read (1988) or Schmitt et al. (2001) as authorities" (J. Read, personal communication, January 28, 2021). One issue with the 26/30 threshold is that 27/30 (90%) is closer to 16/18 (88.8%) than 26/30 (86.6%). The issue with the *use* of the 26/30 mastery threshold in reading research is that reading research indicates that learners need to be able to comprehend 98% of the tokens within a text to easily comprehend it (Schmitt et al., 2011). Furthermore, one purpose for giving levels tests is to match learners with lexically appropriate materials (McLean & Kramer, 2015; Webb et al, 2017), and research suggests that there are several purpose-dependent mastery thresholds.

When matching learners with reading materials through the application of the CCM, the purpose of the reading will determine the most appropriate lexical mastery threshold on levels tests, as well as coverage thresholds when profiling a text. Speed reading involves learners reading materials that contain no unknown words (Nation, 2007). Thus, a mastery threshold of 100% is necessary. Meaning-focused input materials (including extensive reading) require learners to know 98% of the tokens within them (Nation, 2007; Webb & Nation, 2017). Thus, an appropriate threshold is 98%. If the purpose is reading comprehension, then research suggests that an appropriate threshold is 95% (Laufer, 1989; Schmitt et al., 2011). If the purpose for reading is language-focused instruction, Stoeckel et al. (2020) and Schmitt et al. (2011) suggest a threshold no lower than 85%. While the precision of these figures might be questioned, if they are based on research they can be evaluated rationally. More important than the figures themselves, is authors, readers, reviewers, and editors evaluating and justifying the appropriateness of lexical thresholds based on their purpose.

The following section considers the question of why the 86.6% figure is still used, in the hope that the selection of future mastery thresholds will be based on research. First, past practice has not been questioned, perhaps because it makes conducting or publishing research easier, despite the critical examination of previous research being an essential part of the scientific process. Further, by using the 86.6% figure without research-supported justification, subsequent citation of the 86.6% figure becomes easier than would be the case for a research-supported figure.

Second, the consideration of construct validity is all too often absent in vocabulary and extensive reading research methodology, and its application can be labor-intensive and reduces the appearance of robustness. Finally, even if a learner has 98% knowledge of a 1,000-word band, a 98% mastery threshold provides little allowance for measuring errors in test-taker performance, item writing, and the sampling of words from 1,000-word lists. This is particularly problematic for monolingual English vocabulary tests, because writing short definitions of some words using only more frequent words is problematic. Stoeckel et al. (2019) found that learners correctly answered bilingual meaning-recognition items more often than monolingual English meaning-recognition items for the same words that they correctly answered in a meaning-recall format.

Word counting units, and their impact on estimating the lexical load of a text and the lexical mastery level of a learner

Word counting units. In L2 English research, the most often discussed word counting units (WCU) are (a) the *type,* an orthographic form (*use*); (b) the *lemma,* a base word of a particular part of speech (POS) and inflectional forms (use_{verb}, $used_{verb}$, $uses_{verb}$, $using_{verb}$); (c) the *flemma*, a base word form and inflectional forms, regardless of POS (use_{verb}, $used_{verb}$, $used_{adjective}$, $uses_{verb}$, $using_{verb}$, use_{noun}, $uses_{noun}$); (d) and the *Word Family* (WF6), a base word form, inflectional forms, and derivational forms regardless of POS to level 6 of Bauer and Nation's affix criteria (use_{verb}, use_{noun}, $misuse_{verb}$, $misused_{verb}$, $misused_{adjective}$, $misuser_{noun}$, $misusers_{noun}$, $misuses_{verb}$, $misusing_{verb}$, $reusable_{adjective}$, $reuse_{verb}$, $reused_{adjective}$, $reused_{verb}$, $reuses_{verb}$, $reusing_{verb}$, $unusable_{adjective}$, $unused_{verb}$, $unused_{adjective}$, $usability_{noun}$, $usable_{adjective}$, $useable_{adjective}$, $used_{verb}$, $used_{adjective}$, $useful_{adjective}$, $usefully_{adjective}$, $uselessnes_{noun}$, $useless_{adjective}$, $uselessly_{adjective}$, $user_{noun}$, $users_{noun}$, $uses_{verb}$, and $using_{verb}$). It should be stressed that flemmas are not lemmas, flemmas are often wrongly labeled as lemmas, and research and pedagogy will benefit from the accurate labeling of WCUs.

WF6 use assumes that learners who can comprehend the base word form or another WF6 form, can receptively infer the meaning of all WF6 derivational forms with little or no effort (Bauer & Nation, 1993), regardless of the frequency of the derivational form (P. Nation, personal communication, March 22, 2021). Thus, research that considers the validity of different WCUs is concerned with a learner's ability to comprehend a base word form and its associated derivational forms, and the frequency of the derivational form is irrelevant (P. Nation, personal communication, March 22, 2021). If the frequency of a derivational form significantly influences its comprehensibility, it is evidence that derivational forms are learned and comprehended as whole words and *not* through applying affix knowledge to known base word forms or word building. This would, in fact, be evidence against the use of WF6.

The WF6 is dominant in second language (L2) reading research, despite studies suggesting that WF6 is not an appropriate WCU for the majority of EFL learners (see Brown et al., 2020 and Stoeckel, 2020, for reviews). Justification for using the WF6 comes from citing L1 research (see McLean & Kramer, 2015), unsupported opinions in books, past practices that are not supported with evidence (see Dang & Webb, 2014, McLean & Kramer, 2015), or simply that WF6 is commonly used (Dang, 2018; Dang et al., 2017). When L2 English research is used to support the use of WF6, the rationale is that derivational knowledge develops with general English proficiency. Dang et al., (2017, p. 14) state:

> In this way, knowledge of one word family member might help learners acquire other members. This assumption is supported by earlier studies showing that L2 learners' derivational knowledge increases incrementally over time (Mochizuki & Aizawa, 2000; Schmitt & Meara, 1997; Schmitt & Zimmerman, 2002) and that instruction about word parts helps to expand learners' vocabulary knowledge (Schmitt & Meara, 1997; Wei, 2014).

However, just because derivational word knowledge develops over time, it does not mean that the derivational word knowledge possessed by the majority of EFL learners—even the participants in Mochizuki and Aizawa (2000) and Schmitt and Meara (1997)—is sufficient to support the use of WF6. Schmitt and Meara's participants improved their mean affix knowledge from 62% to only 66% of tested affixes over two semesters. Similarly, Mochizuki and Aizawa found that participants correctly answered only 56% of the meaning-recognition affix items. Thus, even studies used to support the use of WF6 (Mochizuki & Aizawa, 2000; Schmitt & Meara, 1997) do *not* actually provide evidence that WF6 is appropriate for learners of similar abilities[4].

WCUs are important because of assumptions involving the ability of English learners to comprehend derivational forms. These assumptions directly affect (a) the coverage that 1,000-word bands provide, and (b) the number of associated inflectional and derivational word forms that are assumed to be comprehensible. Gardner (2007, p. 242) states "[f]urthermore, when corpus-based vocabulary findings are used to inform or support actual language acquisition, there is the additional concern of whether researcher-based conceptualizations of *Word* (i.e., the criteria used to group words, count words, etc.) actually match the psychological realities of Word (i.e., actual knowledge of or about words in the minds of target language users)." Thus, ideally, a WCU would only group base word forms and associated word forms that a learner can comprehend. However, the use of different WCUs among learners inhibits the comparison of results; indeed, one reason Bauer and Nation (1993) created the WF6 was to standardize WCUs to facilitate comparisons of research. Furthermore, determining which derivational forms are known by a learner so that the most appropriate WCU can be selected would require hundreds of thousands of derivational forms to be tested individually.

An alternative approach is to group derivational forms that include known affixes, but this, too, has limitations. First, evidence suggests that the ability to comprehend derivational word forms containing the same affix is base word dependent. For example, in McLean (2018), 276 of the 277 participants, who correctly provided the meaning of the word *teach,* also correctly provided the meaning of the word *teacher*. This might suggest that these 276 participants can comprehend derivational forms containing *-er* provided they can also comprehend the associated base word form. However, of the 268 participants who correctly provided the meaning of the word *develop,* only 225 correctly provided the meaning of *developer*. This calls into question the value of diagnostic affix tests such as the Word Part Levels Test (Sasao & Webb, 2017) because when a learner demonstrates knowledge of an affix in isolation and in a multiple-choice question, it does not necessarily mean they can comprehend word forms composed of the tested affix, even if they can comprehend the base word. Second, evidence suggests that derivational forms containing

[4] Wei (2014) refers to the use of word parts to learn and retain words. Schmitt & Zimmerman (2002) refers to the learners' ability to produce and not comprehend words of different parts of speech.

multiple affixes are more difficult to comprehend. In McLean (2018), of the 277 participants who correctly provided the meaning of the word *use*, 243 correctly provided the meaning of *reuse*, and 221 participants correctly provided the meaning of *usable*. However, only 202 participants correctly provided the meaning of the word *reusable*. Importantly, 26.8% of the derivational forms in Nation's (2006) first five 1,000-WF6 list of the British National Corpus (BNC) include two or more affixes (Brown, 2018). Thus, predicting the difficulty of derivational forms from the difficulty of their affix(es) is problematic.

The use of a single WCU, for example in EFL settings, facilitates comparisons between studies. The choices are the lemma (in practice difficult, as word profilers often do not distinguish between word forms of the same part of speech), the flemma (all too often incorrectly labeled the lemma), or the WF6.

The impact of a word counting unit on estimating the lexical load of texts. Research often establishes which corpus-derived 1,000-word bands provide 98% coverage of various reading materials. If the 1,000-word bands are WF6-based, then it is assumed that learners can comprehend all derivational forms within the WF6, as the occurrence of infrequent derivational forms (e.g., *usage, usability, uselessness*) are counted along with more common derivational forms (*usable, useful*), inflectional forms (*using, used*), and base word form(s) (use_{verb}, use_{noun}). For example, the first 1,000-WF6 in the BNC wordlists (Nation, 2006) consists of 6,857-word types (P. Bennett, personal communication, January 8, 2021). Laufer and Cobb (2020) argue that the ability to comprehend derivational forms containing less frequent affixes is not necessary for reading comprehension, as learners rarely meet them when reading. In contrast, Brown (2018) concluded that derivational forms containing less frequent affixes are important for reading comprehension. Brown (2018) analysed the first five 1,000-WF6 bands of the BNC lists and estimated that texts with 95% coverage based on WF6 have only 82.3, 86.6, 89.5, and 91.2% coverage when knowledge of levels two-, three-, four-, and five-word family forms are included, respectively (Table 1). Similarly, texts with 98% coverage based on WF6 only have 84.9, 89.3, 92.4, and 94.1% coverage if knowledge of levels two-, three-, four-, and five-word family forms are included, respectively (Table 1). This is important because Schmitt et al. (2011) found that a 1% reduction in coverage reduced comprehension by 2.3% (between 92% and 100% coverage).

Table 1

How Assumed Coverage Levels Are Affected by Different Levels of Affix Knowledge

Assumed coverage	If base words only are known	Plus other forms	Plus level 2 forms	Plus level 3 forms	Plus level 4 forms	Plus level 5 forms	Plus level 6 forms
95	58.9	60.1	82.3	86.6	89.5	91.2	95.0
98	60.8	62.0	84.9	89.3	92.4	94.1	98.0

Note. Adapted from "Examining the word family through word lists" by Brown, 2018, p. 59.

The impact of a word counting unit on estimating the lexical mastery level of a learner. The WCU used by levels tests impacts the estimation of a learner's lexical mastery level because each WCU is associated with assumptions about the learner's knowledge of derivational forms. If a levels test adopts the flemma or lemma, correctly answering a word's base word form is interpreted as knowing all associated inflectional forms of the same POS or all associated inflectional forms of any POS (Kremmel, 2016), respectively. For example, if a learner correctly answers 30 items representing the first 1,000 flemmas of the Corpus of Contemporary American English (COCA) (Davies, 2008), the test administrator assumes that the learner knows the first 1,000 flemma or the 3,580-word forms (not including the proper nouns or abbreviations) making up the first 1,000 flemmas (G. Pinchbeck, personal communication, January 7, 2021). In contrast, if a levels test adopts the WF6, correctly answering a word's base word form is interpreted as knowing all associated WF6 inflectional and derivational forms (Kremmel, 2016), or 7,235-word forms (G. Pinchbeck, personal communication, January 7, 2021). If learners can comprehend these additional 3,655 derivational forms when reading, then a learner's knowledge is not overestimated. However, L2 English research strongly suggests that this is not the case.

Studies investigating L2 English learners' ability to comprehend derivational forms. Eight studies provide insight into L2 English learners' receptive knowledge of derivational forms (Brown, 2013; Laufer et al., 2021; McLean, 2018; Mochizuki & Aizawa, 2000; Sasao & Webb, 2017; Schmitt & Meara, 1997; Stoeckel et al., 2018; Ward & Chuenjundaeng, 2009)[5]. None of the seven studies which provide a detailed breakdown of their data, present evidence that all the participants comprehended all the derivational forms that (a) make up WF6, or (b) employ the most common ten affixes as defined by Laufer & Cobb (2020) or Sánchez-Gutiérrez et al. (2018) (Tables 2 and 3). Thus, while claims that such research is *limited* (i.e., Laufer & Cobb, 2020) are correct, by engaging with the research, a balanced inference can be drawn that the use of WF6 is inappropriate in the majority of EFL settings. Table 2 demonstrates that 55.53% (low group) 60.93% (high group) of Thai participants, and 62.3% of Japanese participants were unable to comprehend derivational forms containing six of the most common affixes. Table 3 shows that 67% of Japanese and 86.1% of various L2 English participants demonstrated meaning-recognition knowledge of derivational forms containing the most common affixes. Furthermore, Stoeckel et al. (2018) found that participants were unable to comprehend the meaning of two identical base word forms of different POS 42% of the time.

[5] Laufer et al (2021) found no significant difference for scores from the VST and a custom-made 'Derivatives Test' among advanced learners, while a significant difference was found among less advanced learners. The way in which the data is presented in Laufer et al (2021) means it is not possible to establish if all of the participants comprehended all the derivational forms that (a) make up WF6, or (b) employ the most common ten affixes as defined by Laufer & Cobb (2020) or Sánchez-Gutiérrez et al. (2018).

Table 2

Written Receptive Meaning-recall Knowledge of Derivational Forms Featuring Frequent Affixes (percent correct); after Stoeckel, McLean, and Nation (2020)

	Ward & Chuenjundaeng (2009)		McLean (2018)			
Affix	Participants		Participants			
	Low group	High group	All	Beginner	Intermediate	Advanced
-ly						
-ion	58.5	31.1				
-er	66.9	94.2				
-y						
-al			84.5	79.8	86.4	88.2
re-			79.7	66.5	83.7	98.8
un-**						
-age**			22.7	9.5	25.0	64.7
-ness**						
-ity	41.2	57.5				
-ate*						
-in*						
-ant*						
Mean	55.5	60.9	62.3	52.0	65.0	83.9

Note. * Affixes that Sánchez-Gutiérrez et al. (2018) identified as being among the ten most common affixes of English. ** Affixes that Laufer and Cobb (2020) identified as among the ten most common affixes of English. All other affixes were identified as being among the ten most common affixes of English by both Sánchez-Gutiérrez et al. (2018) and Laufer and Cobb (2020).

Table 3

Recognition of Affix Meaning and Affix Grammatical Function for Frequent Affixes (percent correct); after Brown, Stoeckel, McLean, and Stewart (2020)

Affix	Mochizuki & Aizawa (2000)		Sasao & Webb (2017)[a]	
	Meaning recognition	Grammatical function recognition	Meaning recognition	Grammatical function recognition
-ly				47.2 (adjective use) 69.1 (adverb use)
-ion				64.2
-er		75	94.8	89.4
-y		42		70.6 (adjective use) 44.2 (adverb use)
-al		75		68.2 (adjective use) 60.9 (noun use)
re-	93		94.0	
un-**	81		88.2	
-age**				55.2
-ness**		67		59.6
-ity		34		52.8
-ate*				61.0
in-*	27		60.0	
-ant*			91.1	60.7 (adjective use) 85.1 (noun use)
Mean	67	58.6	86.1	63.59

Note. * Affixes that Sánchez-Gutiérrez et al (2018) identified as being among the ten most common affixes of English. ** Affixes that Laufer and Cobb (2020) identified as being among the ten most common affixes of English. All other affixes were identified as being among the ten most common affixes of English by both Sánchez-Gutiérrez et al. (2018) and Laufer and Cobb (2020). The figures in these columns include data for only L2 English users.

Underestimation of derivational knowledge is preferable to its overestimation. In defense of WF6, the adoption of the flemma or lemma will underestimate all learners' ability to comprehend some derivational forms. Table 2 suggests that around 60% of participants can recall the meanings of tested forms containing high frequency affixes. Thus, the use of the flemma or lemma would understate the derivational knowledge of these participants. However, there are four reasons why overestimating derivational word knowledge is of greater concern than the opposite when operationalizing the CCM. First, as previously stated, a 1% decline in coverage reduces comprehension by 2.3% (Schmitt et al., 2011) (between 92% and 100% coverage). Thus, even a slight overestimation of comprehension has a considerable impact. Second, overestimating a learner's ability to comprehend derivational forms can result in the use of texts that are incomprehensible and/or lexically too difficult for their purpose. Third, the coverage window for unassisted comprehension is only 5%—between 95% and 100% coverage. Thus, any overestimation of derivational knowledge can quickly result in an incomprehensible text. In contrast, the underestimation of derivational knowledge and coverage will result in greater, or full, coverage, something that most people experience when reading in the L1 and which is still beneficial for L2 reading development. Finally, lexical coverage figures for texts and a learner's lexical mastery estimate assume that all proper nouns and homoforms are known by learners. However, research suggests that this is not the case (Brown, 2010; Klassen, 2018). Thus, even if a learner is expected to know 100% of the tokens in a text from the application of the CCM, homoforms and proper nouns can reduce or inhibit comprehension of a text. Thus, this paper recommends using the lemma or flemma, and not WF6, as a general WCU in EFL settings. The adoption of the lemma, however, will require the creation of easily usable lexical profilers that identify a word's POS, so until then, the flemma is a pragmatic choice.

Solutions to the limited validity with which the text coverage model is operationalized

The CCM is the cornerstone of reading/lexical research. However, the limited validity with which research operationalizes the CCM is the result of assumptions that research suggests are incorrect. There are two simple ways to improve the appropriateness of lexical coverage investigations. First, researchers can select a WCU based on existing research or conduct their own investigations to demonstrate that the WCU used in their research is appropriate. Presently, the dominance of WF6 is despite the published evidence and not because of it. Thus, it is hoped that reviewers and editors require an empirical, research-based justification for researchers' choice of WCU, and that reviewers and editors critically examine the justification. Stating that a given WCU was used in past research or that it is commonly used is *not* evidence of its appropriateness. Similarly, simply stating the proficiency of the target learners is *not* evidence that a WCU is appropriate for a given purpose. Second, coverage thresholds and levels test mastery thresholds should be evidence based and set per the purpose for reading.

As a final thought, it might not be possible to operationalize the CCM with a high degree of validity owing to natural measurement error. It might be the case that the CCM places too much importance upon vocabulary, and the CCM in practice does not allow teachers and researchers to predict if texts can be comprehended or not. However, unless the CCM is first correctly operationalized, validity cannot be ascertained.

Acknowledgement

The author would like to thank Dale Brown, Tim Stoeckel, Phil Bennett, Christopher Nicklin, Jeff Stewart, and Steve Porritt for their feedback on this article. Any remaining faults are the sole responsibility of the author.

References

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Brown, D. (2010). An improper assumption? The treatment of proper nouns in text coverage counts. *Reading in a Foreign Language, 22*(2), 355–361. http://hdl.handle.net/10125/66842

Brown, D. (2013). Types of words identified as unknown by L2 learners when reading. *System*, *41*(4), 1043–1055. https://doi.org/10.1016/j.system.2013.10.013

Brown, D. (2018). Examining the word family through word lists. *Vocabulary Learning and Instruction, 7*(1), 51–65. https://doi.org/10.7820/vli.v07.1.brown

Brown, D., Stoeckel, T., Mclean, S., & Stewart, J. (2020). The most appropriate lexical unit for L2 vocabulary research and pedagogy: A brief review of the evidence. *Applied Linguistics*. https://doi.org/10.1093/applin/amaa061

Dang, T. N. Y. (2018). A hard science spoken word list. ITL - *International Journal of Applied Linguistics*, *169*(1), 44-71. https://doi.org/10.1075/itl.00006.dan

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning, 67*(4), 959–997. https://doi.org/10.1111/lang.12253

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes, 33*, 66–76. https://doi.org/10.1016/j.esp.2013.08.001

Davies, Mark. (2008) *The corpus of contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition, 42*(3), 499–523. https://doi.org/10.1017/S0272263119000494

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, *28*(2), 241–265. https://doi.org/10.1093/applin/amm010

Gui, M., Shang, Y., & Chen, X. (2020). Effect of timed reading on Chinese undergraduates' EFL reading rates: Mixed-method analyses. *Reading in a Foreign Language, 32*(2), 104–121. http://hdl.handle.net/10125/67376

Gyllstad, H., McLean, S., & Stewart, J. (2020). Using confidence intervals to determine adequate item sample sizes for vocabulary tests: An essential but overlooked practice. *Language Testing*, 0265532220979562. https://doi.org/10.1177/0265532220979562

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

Hu, H. M., & Nation, P. (2000). What vocabulary size is needed to read unsimplified texts. *Reading in a Foreign Language*, *8*, 689–696. http://hdl.handle.net/10125/67046

Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language*, *26*(2), 17–33.

Klassen, K. (2018). *Investigating the lexical load of proper names for L2 English readers* (Doctoral dissertation, Cardiff University).

Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly, 50*(4), 976–987. https://doi.org/10.1002/tesq.3

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Laurén & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.

Laufer, B., & Cobb, T. (2020). How much knowledge of derived words is needed for reading? *Applied Linguistics*, *41*(6), 971–998. https://doi.org/10.1093/applin/amz051

Laufer, B., Webb, S., Kim, S. K., & Yohanan, B. (2021). How well do learners know derived words in a second language? The effect of proficiency, word frequency and type of affix. *ITL-International Journal of Applied Linguistics*, Online-First Articles. https://doi.org/10.1075/itl.20020.lau

McLean, S. (2014). Evaluation of the cognitive and affective advantages of the Foundations Reading Library series. *Journal of Extensive Reading*, *2*, 1–14.

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, *39*(6), 823–845. https://doi.org/10.1093/applin/amw050

McLean, S., & Kramer, B. (2015). The creation of a New Vocabulary Levels Test. *Shiken, 19*(2), 1–11.

McLean, S., Stewart, J., & Batty, A. O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing*, *37*(3) 389–411. https://doi.org/10.1177/0265532219898380

Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, *28*(2), 291–304. https://doi.org/10.1016/S0346-251X(00)00013-0

Nation, P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language, 26*(2), 1–16. http://hdl.handle.net/10125/66881

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

Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching, 1*(1), 2–13. https://doi.org/10.2167/illt039.0

Nation, P. (2016). *Making and using word lists for language learning and testing*. John Benjamins Publishing Company. https://doi.org/10.1075/z.208

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal 19*, 12–25. https://doi.org/10.1177/003368828801900202

Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. H., & Wilson, M. A. (2018). MorphoLex: A derivational morphological database for 70,000 English words. *Behavior Research Methods*, *50*(4), 1568–1580. https://doi.org/10.3758/s13428-017-0981-8

Sasao, Y., & Webb, S. (2017). The word part levels test. *Language Teaching Research*, *21*(1), 12-30. https://doi.org/10.1177/1362168815586083

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

Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching, 53*, 109–120. https://doi.org/10.1017/S0261444819000326

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

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36*(2), 145–171. https://doi.org/10.2307/3588328

Stoeckel, T., Ishii, T., & Bennett, P. (2018). Is the lemma more appropriate than the flemma as a word counting unit? *Applied Linguistics*, *41*(4), 601–606. https://doi.org/10.1093/applin/amy059

Stoeckel, T., McLean, S., & Nation, P. (2020). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 1–23. https://doi.org/10.1017/S027226312000025X

Stoeckel, T., Stewart, J., McLean, S., Ishii, T., Kramer, B., & Matsumoto, Y. (2019). The relationship of four variants of the Vocabulary Size Test to a criterion measure of meaning recall vocabulary knowledge. *System*, *87*, 102–161. https://doi.org/10.1016/j.system.2019.102161

Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, *35*(2), 182–191. https://doi.org/10.1016/j.system.2006.12.009.

Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, *37*(3), 461–469. https://doi.org/10.1016/j.system.2009.01.004

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*, 130–163. http://hdl.handle.net/10125/66776

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics, 168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

Wei, Z. (2014). Does teaching mnemonics for vocabulary learning make difference? Putting the keyword method and the word part technique to the test. *Language Teaching Research, 19*, 43–69. https://doi.org/10.1177/1362168814541734

Zhang, S., & Zhang, X. (2020). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research, OnlineFirst*. https://doi.org/10.1177/1362168820913998

About the Author

Stuart McLean is interested in vocabulary and reading research, and the importance of construct measurement within research design. He is currently making online self-marking form-recall and meaning-recall (orthographic and phonological) vocabulary levels tests, that allow teachers to create levels tests based on various (a) lists, (b) word-band sizes, (c) band ranges, and (d) sampling ratios. Teachers can download automatically marked responses, actually typed responses, and the time taken to complete responses. Presently tests are designed for Japanese learners studying English, and English speakers learning Spanish (vocableveltest.org).
E-mail: stumc93@gmail.com