
Article:
González-Fernández, B. orcid.org/0000-0002-4370-1822 (2022) Conceptualizing L2 vocabulary knowledge. Studies in Second Language Acquisition, 44 (4). pp. 1124-1154. ISSN 0272-2631

https://doi.org/10.1017/s0272263121000930

Conceptualizing L2 Vocabulary Knowledge: An Empirical Examination of the Dimensionality of Word Knowledge

Abstract

The multidimensional conceptualisation of vocabulary knowledge has been extraordinarily influential in lexical research. Yet, the few studies that have empirically explored the structure of word knowledge conflict regarding its multidimensionality. The present study examines the nature of L2 vocabulary knowledge by exploring how the hypothesised word-knowledge dimensions fit together across different L2 learner groups. A total of 314 EFL learners from two dissimilar L1 backgrounds (Chinese n = 170; Spanish n = 144) were assessed via eight vocabulary measures testing recognition and recall knowledge of form-meaning, collocations, multiple-meanings and derivatives. CFA revealed that vocabulary knowledge behaves as a unidimensional construct for each learner group individually. Simultaneous multigroup CFA confirmed that this unidimensional model holds invariant across the two groups regardless of their distinct L1s. These findings provide empirical evidence for the unidimensionality of vocabulary knowledge in a second language, indicating the need to refine the conceptualisation of the construct.

Introduction

The vocabulary-knowledge construct is theoretically described, interpreted and accepted as multidimensional by second language acquisition (SLA) researchers and practitioners. This multifaceted description acknowledges the general consensus among lexical researchers that knowing vocabulary is not an all-or-nothing construct, but rather it involves mastering multiple and independent dimensions of knowledge for each lexical item, for example, a word's spelling, its morphological forms, its multiple meanings depending on the context, or the words with which it can collocate (Nation, 2020; Schmitt, 2014; Webb, 2005).

Despite this agreement on the theoretical multidimensionality of vocabulary knowledge, research has rarely explored how these hypothesised dimensions of word knowledge relate to each other and whether they can indeed be reliably and empirically distinguished as independent dimensions, which would mean that the components develop as separate entities (Kieffer & Lesaux, 2012). For example, interpreting that the form-meaning link and the spoken form of words represent two distinct lexical dimensions implies that, while they can be interrelated to some extent, one can grow with relative independence from the other. Thus, this multidimensional conceptualisation, which has influenced pedagogy and research for decades, remains a hypothesised description of the nature of vocabulary knowledge. Understanding how well this theorisation represents learners' vocabulary knowledge in real life is essential in order to provide validity to the lexical descriptions and to inform second language (L2) theory, research, assessment and teaching (Milton & Fitzpatrick, 2014; Schmitt, 2019).

Consequently, a number of SLA scholars (e.g., Godfroid, 2020; Kieffer & Lesaux, 2012; Nation, 2020) have called for research which investigates how the various hypothesised dimensions of word knowledge are known and fit together in constructing the general structure of vocabulary knowledge. In an initial effort to explore this matter, González-Fernández and

Schmitt (2020) investigated Spanish learners of English and found that, contrary to the widely-accepted multidimensional descriptions, vocabulary knowledge behaved as a unidimensional construct[1] for this group of L2 learners. This unidimensionality means that the various types of word knowledge are highly integrated aspects of lexical knowledge that function inseparably from each other. Despite this preliminary evidence of the unidimensionality of vocabulary knowledge, the great complexity of the construct and the potential influence of learners' L1 on their lexical knowledge demand further investigation to untangle and validate the word-knowledge nature across different types of L2 learners.

The goal of this study is to begin to empirically model the dimensional structure of vocabulary knowledge in second languages by examining the relationships among word-knowledge dimensions in the lexical competence of English-as-a-foreign-language learners (EFL) from different L1 backgrounds. Employing Nation's (2013) multidimensional word-knowledge framework (see following section), the study tests EFL learners from two unrelated L1s (i.e., one cognate and one non-cognate) in order to better understand how well the multidimensional conceptualisation of vocabulary knowledge represents the actual structure of L2 word knowledge when compared to a unidimensional model. It also discusses the influence that measuring various types of word knowledge of the same or distinct sets of items has on the examination of the nature of vocabulary knowledge. If the various word-knowledge dimensions are found to behave as separate and distinct knowledge entities, this would validate the currently-accepted conceptualisation of vocabulary knowledge as multidimensional. However, if the various word-knowledge dimensions are found to be known inseparably from each other across the two learner groups, this would suggest that L2 vocabulary knowledge is unidimensional, and that the word-knowledge components cannot be identified as *dimensions*. A unidimensional result would require a reconceptualization of the nature of vocabulary knowledge, with important implications for both theory and practice.

Theoretical conceptualisations of vocabulary knowledge

In an attempt to explain the complex and elusive nature of vocabulary knowledge, researchers have proposed several and varied theoretical conceptualisations of the construct (e.g., Cronbach, 1942; Nation, 2013; Qian, 2002). These conceptualisations have typically been categorised into two main approaches: developmental approach and components or dimensions approach (Read, 2000). The *developmental approach* understands vocabulary knowledge as existing along a scale or continuum, where the different types of word knowledge (form recognition, meaning recall, etc.) are organised by difficulty into hypothesised hierarchical scales, from no knowledge at all to full proficiency of a word (e.g., Wesche & Paribakht, 1996). Despite its practicality in research, this approach presents several limitations as a theoretical description of word knowledge. These include, but are not limited to, its partial account of the various aspects involved in vocabulary knowledge, or the fact that the hypothesised hierarchical scales of word-knowledge types have not been empirically validated (Stewart et al., 2012), making it unclear whether word knowledge develops in the order the scales propose (see Yanagisawa & Webb, 2020 for a detailed review). These limitations mean that the developmental conceptualisation of vocabulary knowledge has not been widely embraced in lexical research.

In contrast, the *dimensions* or *components approach* has exerted a great influence on vocabulary studies. This approach divides up the vocabulary-knowledge construct into multiple word-knowledge components or dimensions. Since Cronbach's (1942) early description of five word-knowledge dimensions, which include general components such as size, depth and automaticity, researchers have continued to propose different classifications and identify more precise dimensions of vocabulary knowledge. The most complete and refined list to date has been proposed by Nation (2013), who suggests that knowledge of a word involves achieving proficiency in nine aspects, each divided into receptive and productive

ability (Figure 1). This detailed description of word-knowledge types provides a rich and precise picture of the multidimensionality of lexical knowledge, which has made Nation's framework the preferred and most widely-accepted conceptualisation among researchers investigating L2 vocabulary knowledge (e.g., Cheng & Matthews, 2018; Li & Kirby, 2015; Webb, 2005).

Figure 1 Nation's (2013, p. 49) framework of word-knowledge dimensions

FORM	Spoken	[R]	What does the word sound like?
		[P]	How is the word pronounced?
	Written	[R]	What does the word look like?
		[P]	How is the word written and spelled?
	Word parts	[R]	What parts are recognisable in this word?
		[P]	What word parts are needed to express the meaning?
MEANING	Form and meaning	[R]	What meaning does this word form signal?
		[P]	What word form can be used to express this meaning?
	Concept and referents	[R]	What is included in the concept?
		[P]	What items can the concept refer to?
	Associations	[R]	What other words does this make us think of?
		[P]	What other words could we use instead of this one?
USE	Grammatical functions	[R]	In what patterns does the word occur?
		[P]	In what patterns must we use this word?
	Collocations	[R]	What words or types of words occur with this one?
		[P]	What words or types of words must we use with this one?
	Constraints on use	[R]	Where, when and how often would we expect to meet this word?
		[P]	Where, when, and how often can we use this word?

Both the developmental and the components approaches describe vocabulary knowledge as the mastery of various types of knowledge for each lexical item; thus, understanding it as a multidimensional construct. Yet, despite their substantial influence in L2 vocabulary theory, research and pedagogy, these theorisations have typically not been empirically examined and validated in research. Nation (2020) recently acknowledged that his detailed word-knowledge framework, while being a convenient description of the possible range of knowledge dimensions learners can have about words, cannot explain how these hypothesised dimensions behave in actual vocabulary use, relate to each other and contribute to the overall vocabulary-knowledge structure. Therefore, there is a need to investigate the empirical dimensionality of L2 vocabulary knowledge in order to establish the relationships

between these knowledge dimensions and provide validity for the theoretical descriptions of the construct (Kieffer & Lesaux, 2012; Spencer et al., 2015; Stewart et al., 2012). Specifically, given its extensive use in vocabulary acquisition research, it is essential that Nation's word-knowledge framework and the multidimensional conceptualisation it represents is validated. This validation would provide an empirically-supported model of L2 vocabulary knowledge that allows lexical research results to be interpreted with confidence.

Relationships among vocabulary components

While research examining how the various word-knowledge dimensions fit together is still limited, there is a general consensus among researchers that these dimensions are intrinsically interconnected and interact with each other (Li & Kirby, 2015; Milton & Fitzpatrick, 2014). For example, strong correlations have been found between knowledge of the spoken and written form of words (r = .59-.81, Cheng & Matthews, 2018). Similarly, vocabulary size (i.e., knowing the form–meaning link of words) has been found to correlate moderately to highly with multiple word-knowledge components, including derivatives (a type of word-parts knowledge in Nation's framework) (r = .54-.65, Mochizuki & Aizawa, 2000), associations (r = .70-.82, Qian, 2002), and collocations (r = .83-.88, Gyllstad, 2009; r =.70, Nguyen & Webb, 2017). The productive use of words also correlates strongly with word class (r = .62), meaning (r = .50), derivatives (r = .71) and collocational knowledge (r = .75) (Chui, 2006). Likewise, the recall and recognition aspects[2] of form-meaning link, derivatives, multiple meanings and collocations have exhibited strong correlations with each other (r =.70-.95, González-Fernández & Schmitt, 2020).

These studies share methodological features that allow them to inform about the interconnectedness among word-knowledge components. They conceptualise word knowledge following Nation's (2013) framework, adopt similar recall and recognition formats to assess

the components, measure real words (except Mochizuki & Aizawa), and control for the order of tests administration. When the studies assess vocabulary size, they typically adapt standardised tests, and thus assess different words across tasks. However, when the studies measure various depth-of-knowledge aspects, they tend to test the same target words across tasks (except Chen & Matthews).

Overall, the available evidence indicates that the various components involved in vocabulary knowledge are positively and highly interconnected with each other (Qian, 2002; Schmitt, 2014; Webb, 2005), suggesting that no component is known in a completely detached manner from the other components. These high intercorrelations among the word-knowledge components (specifically between the form-meaning link and depth[3] components) have led some researchers to question whether they are indeed distinct dimensions, with some scholars even suggesting that they might be better seen as the same construct (e.g., Gyllstad, 2013; Nguyen & Webb, 2017; Yamashita & Jiang, 2010). In light of these findings, research is needed to empirically validate the hypothesised multidimensionality of L2 vocabulary knowledge.

The nature of vocabulary knowledge

Research that has investigated the presupposed multidimensionality of vocabulary knowledge empirically is scarce (Kieffer & Lesaux, 2012; Li & Kirby, 2015), and thus the dimensional nature of vocabulary knowledge has not yet been verified. This section reviews the handful of studies that have attempted to address this issue.

One of the first attempts to inform about the dimensionality of vocabulary knowledge is Kieffer and Lesaux (2012). They studied the vocabulary knowledge of sixth-grade students in the US with L1 and L2 English (from various first languages, with Spanish accounting for 73% of the L2 sample). Following a components approach, seven reading-based tests were administered, measuring receptive knowledge of various lexical components across different

sets of words. Using multigroup confirmatory factor analyses, they found that a three-dimensional model of vocabulary knowledge consisting of *breadth* (represented by academic synonyms, general synonyms, multiple meanings and semantic associations tasks), *contextual sensitivity* (informed by a guessing-from-context test) and *morphological awareness* (assessed by real-word decomposition and nonword derivation measures), was the best-fitting model for their L1 and L2 English participants. The authors concluded that a three-dimensional model comprised of three highly connected (r = .71-.85) but distinct dimensions was the best representation of vocabulary knowledge. Their multidimensional conceptualisation of lexical knowledge was operationalised via some measures that match Nation's (2013) components (e.g., semantic associations, multiple meanings [Nation's concepts and referents], or real-word decomposition [word parts]). Yet, the study also assessed linguistic skills that go beyond this framework (i.e., guessing from context). This approach could influence the dimensionality results by bearing additional dimensions that would not be featured in studies which strictly adopt Nation's framework as their underlying conceptualisation (e.g., González-Fernández & Schmitt, 2020).

In a research paper with only L1-English children, Spencer et al. (2015) examined the underlying dimensions of morphological awareness and vocabulary knowledge based on the hypothesis that they are two related but independent constructs. The paper comprised two studies with fourth-graders and eight-graders, respectively. Morphological awareness was operationalised in both studies using a series of tests tapping into various types of morphological knowledge (e.g., derivations, inflections and compounds). All tasks, except one in the first study, tested real words via a variety of formats and administration modes, including written recall of inflectional and derivational words, oral word decomposition and derivation, and written and oral multiple-choice recognition of morphemes and suffixes. Vocabulary knowledge was assessed in Study 1 by means of size tests measuring meaning recall and

meaning recognition. In the second study, vocabulary was understood more holistically as knowledge of size and depth to try and capture various dimensions of word knowledge, in a manner more compliant with Nation's (2013) framework; thus, vocabulary was tested with a meaning-recall task measuring size, a task involving the use of words in context (i.e., grammatical functions, collocations and register) and a synonyms and antonyms test (associations in Nation's terms). In the first study, the various measures tested different target words, while the second study assessed an identical set of 23 words across all the tasks to control for the differential knowledge of specific vocabulary items (e.g., intra-word learning burden, learning experience, etc.). Contrary to their hypothesis, the authors found that a unidimensional model where vocabulary and morphological awareness comprised a single construct provided the best fit to the data in both studies. It was then concluded that all their morphology and vocabulary tasks measured the same underlying ability across different and identical word-sets, suggesting that vocabulary knowledge overall behaves as a unidimensional construct in L1 English. Compared to Kieffer and Lesaux's research, Spencer et al.'s view of vocabulary knowledge (particularly in Study 2) is more congruent with Nation's framework, which might partially explain the conflicting dimensionality findings.

In the L2 context, González-Fernández and Schmitt (2020) examined the dimensional nature of word knowledge in Spanish EFL learners in an initial effort to understand the structure of the EFL vocabulary-knowledge construct. The study measured the same set of 20 words across eight types of word knowledge. As in Spencer et al.'s (2015) study, the authors found that a unidimensional model, where all the tasks loaded into a unique construct, provided the best fit to the data when compared to a multidimensional model. The authors concluded that, contrary to the theoretical descriptions, L2 vocabulary knowledge seems to be a unidimensional construct empirically. However, this conclusion is tentative insofar that the study examined a group of EFL learners from only one L1 background. The learners' L1 is one

of the factors that most influences the knowledge and acquisition of L2 words, with some words being easier to acquire because they are semantically, orthographically and/or phonologically similar to an L1 equivalent (i.e., cognate words) (Vidal, 2011). In this sense, L1 speakers of etymologically-related languages which share a substantial number of cognate words with English (e.g., Spanish) can find these words easier to learn and remember than L1 speakers from non-cognate languages (e.g., Chinese), at least at the receptive level of mastery, which, in turn, might influence their overall L2 vocabulary knowledge (Chen et al., 2012). This difference in L2 acquisition and knowledge between EFL learners from cognate and non-cognate languages may also affect the way the vocabulary-knowledge components interrelate, and, thus, the dimensionality of the construct. Consequently, it is unclear whether this unidimensional result found for one cognate L1 group of EFL learners would also represent lexical knowledge for other EFL learner populations, particularly from non-cognate languages.

In a related study, Koizumi and In'nami (2020) investigated the factor structure of vocabulary knowledge, understood as the combination of size (form-meaning link) and depth (polysemy, collocation and word association) of knowledge, in Japanese L2 English learners. As in Kieffer and Lesaux (2012), the study employed ready-available written tests designed for a different research purpose, and thus, each task measured distinct target words across each word-knowledge component. The authors found that a two-factor model, where size and depth were considered separate but correlated dimensions, fit the data better than a single-factor model for their L2 learners. Given that their participants belonged to a non-cognate L1, unlike in González-Fernández and Schmitt's study, this finding suggests that the learners' L1 may affect the way L2 vocabulary knowledge is structured in practice. Importantly, however, in this study the size and depth factors were shown to correlate extremely highly with each other (r = .94-.95), which raises questions as to how distinct these components actually are and whether the two-dimensional result is empirically well-substantiated or might be an artifact of the

characteristics of the measures employed, in particular the measures' original purpose and the features of the different items assessed in each task.

Overall, the above review shows that there is no consensus about or enough evidence to establish the nature of vocabulary knowledge as either multi- or unidimensional. Furthermore, the conflicting findings by Koizumi and In'nami and González-Fernández and Schmitt with two different language groups suggest that EFL learners' distinct L1 might influence the dimensional structure of L2 vocabulary knowledge. Therefore, there is a need for further empirical investigation into the nature of vocabulary knowledge to determine how the hypothesised dimensions of word knowledge fit together across various L2 groups from different L1s. The present study attempts to shed some light on this issue by examining the dimensionality of L2 vocabulary knowledge across EFL learners from two dissimilar L1 backgrounds (one cognate and one non-cognate), and establishing which components, if any, are indeed psychometrically distinct dimensions (Kieffer & Lesaux, 2012; Li & Kirby, 2015; Stewart et al., 2012).

The current study

This study extends González-Fernández and Schmitt's (2020) paper by investigating an additional L1 population and performing simultaneous multigroup model comparisons across EFL learners. It expands on the findings and limitations of previous research in four main ways. First, previous studies on the topic have assessed vocabulary knowledge by means of ready-available and/or independent tests designed for a different purpose (i.e., Kieffer & Lesaux, 2012; Koizumi & In'nami, 2020). This means that they do not follow one particular vocabulary conceptualisation (e.g., Nation's model) as the common rationale for the selection of each target knowledge dimension and measurement instrument. This combination of frameworks and tasks could result in the overdimensionalisation of vocabulary knowledge, that is, finding

extra dimensions that should not be differentiated within 'word knowledge' as is described in most frameworks. The current study addresses this by adopting Nation's (2013) framework as the underlying conceptualisation of L2 vocabulary knowledge to inform the description of dimensions and the selection of measurement instruments. Under this conceptualisation, vocabulary knowledge is understood as knowledge of multiple word-knowledge aspects for each lexical item. Second, and related to the prior issue, some of these previous studies have employed vocabulary measures that tested different sets of words across the various tasks (e.g., Kieffer & Lesaux, 2012; Koizumi & In'nami, 2020), which might have affected the results about the dimensional structure of L2 vocabulary knowledge. On the contrary, the current study controls for the effect that distinct item sets can have on the dimensionality of vocabulary, and assesses various types of word knowledge about the same set of target words in accordance with Nation's framework. Third, some of the previous studies only measure recognition knowledge (Kieffer & Lesaux, 2012; Koizumi & In'nami, 2020), thus providing a limited account of overall lexical knowledge. The present study gauges knowledge of a wider range of vocabulary aspects, including not only recognition, but also recall mastery. Finally, unlike most previous studies, the current paper not only examined the vocabulary structure that best fits a single group of learners. It also simultaneously investigates two learner groups from different L1 backgrounds (Spanish and Chinese), and employs multigroup confirmatory factor analyses to inspect whether there exists a best-fitting model of vocabulary knowledge that is generalisable across EFL learners from cognate and non-cognate L1s. This approach offers a broader perspective on the dimensionality of vocabulary knowledge in second languages.

Based on this prior research, the validity of two competing vocabulary-knowledge models (namely, a multidimensional and a unidimensional model) will be tested in this study. The following research questions will be addressed:

1. To what extent is a unidimensional or a multidimensional model a better-fitting structure of the vocabulary knowledge of a group of EFL learners from a non-cognate background (i.e., L1-Chinese learners)?
2. To what extent does the same word-knowledge model hold across EFL learners from a non-cognate and a cognate L1 backgrounds, i.e., Chinese and Spanish? Which structure better represents vocabulary knowledge in L2 English?

Method

Sample

The participants in this study consisted of 314 EFL learners (254 females, 60 males), whose age ranged from 18 to 65 years (M = 22.42, SD = 6.19). They belonged to two different L1 backgrounds: 144 were L1-Spanish learners (45.9%), and 170 (54.1%) were L1-Chinese learners. They were recruited as volunteers by the author in Spain, China and the UK, and comprised undergraduate and postgraduate students, primarily from Languages or Education degrees (which might explain the larger number of females), as well as professionals in various fields. They had an academic learning history in L2 English of between 1 and 26 years (M = 11.81, SD = 3.90), and most of them (76.8%), at the time of the study, were studying English formally in different settings, including EFL language classrooms and English as medium of instruction. The dataset includes a representative sample of learners with a range of proficiency levels in English, from beginners to advanced, across the two L1 backgrounds. More than half of the participants (56.4%) reported themselves as having an intermediate general proficiency in English, just above a fifth (22.6%) rated themselves as beginners, and 21% considered themselves advanced users. An objective measure of the participants' estimated vocabulary level in English was collected through a compound score of the 2,000 (2K), 3,000 (3K), 5,000

(5K), and 10,000 (10K) sections of the Vocabulary Levels Test (VLT) (Schmitt et al., 2001). The two subsamples differed somewhat in their knowledge per frequency band (L1-Chinese: 2K = 92.4%, 3K = 74.0%, 5K = 63.9%, 10K = 16.4; L1-Spanish: 2K = 82.9%, 3K = 74.3%, 5K = 71.3%, 10K = 42.9%), but demonstrated a similar general vocabulary level on average (61.7% overall for the L1-Chinese participants and 68% for the L1-Spanish sample).

Measures

González-Fernández and Schmitt's (2020) battery of eight vocabulary measures was administered in order to assess learners' knowledge of multiple word-knowledge dimensions. This comprehensive test battery is based on Nation's (2013) dominant framework of L2 word-knowledge aspects, which understands vocabulary knowledge in the conventional manner as mastery of various knowledge dimensions for each individual word.

This measurement instrument assesses written recognition and recall knowledge of the same set of 20 words across four different vocabulary-knowledge components: form-meaning link, derivatives, collocation, and multiple meanings. Given the impracticality of measuring *all* the aspects specified in Nation's framework, these four components offer a representation of each of the main three constructs of vocabulary knowledge: form, meaning and use. The 20 target words (Appendix 1 in the On-line Supplementary Materials) provide the greatest opportunity to concurrently test the four word-knowledge components by meeting the following criteria: a) range of frequencies (1K-9K), to account for the various proficiencies of participants; b) multiple meanings, with at least three senses as different from each other as possible; c) at least three derivative forms for one of the meaning senses; and d) different parts of speech, to have a representative list of words.

Measuring various word-knowledge components across the same set of items is the standard and preferred method employed by researchers exploring depth of vocabulary

knowledge (e.g., Chui, 2006; Peters, 2016; Webb, 2005). The advantages of this approach are threefold: 1) it allows us to investigate how much a learner may know about each particular word, and thus, to examine overall word knowledge as described by most multidimensional frameworks (e.g., Nation, 2020); 2) it provides more meaningful, interpretable and directly comparable scores than testing different words across dimensions, since it ensures that each dimension is being measured in a consistent and balanced manner in terms of the characteristics of the target words; 3) it controls for the potential effect that differential knowledge and features of the words assessed in each measure may have on the dimensionality of vocabulary knowledge (Spencer et al., 2015).

This approach, however, implies showing the specific set of words in multiple contexts and situations, and thus some practice effect between tests may occur. The test battery was purposefully designed to minimise the influence of this effect and so that individual tests did not overlap. Actions taken to this aim included giving the recall tests before their recognition counterpart, making sure that the targeted derivatives, senses and collocations did not appear in the test before the section in which they were measured, and thus could not be answered based on information from previous sections, and inserting parts of the Vocabulary Levels Test between some sections to minimise any possible memory of the target words from previous tests. Nevertheless, to estimate the potential existence of practice effect, the current study conducted post-data-collection examinations. Cochran's Q test with McNemar's post-hoc analyses was computed to identify the proportion of participants who achieved mastery (established at 75% and 80% task accuracy) of typically more complex and later tests (e.g., derivatives or multiple-meanings (Chui, 2006; Schmitt, 2014)) without having mastered the easier aspects (e.g., form-meaning recognition), which would indicate practice effect. The results revealed that there was a statistically significant difference in the proportion of participants who mastered more difficult aspects in the battery without having mastered easier

ones, as compared to the opposite scenario ($\chi^2(7) = 856.51$, $p = .000$). Specifically, very few participants showed mastery of a task without having demonstrated mastery of a generally easier and better-known task (e.g., form-meaning recognition): on average 3.49% participants with the 80% accuracy level and 3.58% with the 75% accuracy level (see Appendix 2 for the tasks' mean scores, which demonstrate the same pattern of mastery in each language group). These results seem to confirm that the influence from measuring the same target words across multiple aspects in this battery was minimal. Therefore, although some degree of practice effect in repeated testing circumstances is probably unavoidable (Nation & Webb, 2011), the analysis provides evidence that the efforts to control for and keep to a minimum the potential practice effect and overlap in this battery were largely successful, and thus the results in this study have generally not been influenced by this effect.

The test battery (available as On-line Supplementary Materials) exhibited high construct reliability for the participants in the present study (Composite Reliability = .94 for the L1-Chinese participants, .98 for L1-Spanish participants, and .96 for the total sample), confirming that the various tests tapped into overall vocabulary knowledge appropriately. The individual tests are briefly described below, and a sample item for each is provided. A more detailed description of each measure, their scoring methods, and the rationale for their selection is offered in the online supplementary materials of González-Fernández and Schmitt (2020) and openly available in the IRIS digital repository (https://www.iris-database.org).

Form recall knowledge of the form–meaning mapping

This test employed a fill-in-the-blank format where participants were asked to recall the English form of a word, given its most frequent meaning in the participants L1. The meaning of the target item was provided in context, as in the example below (translation: "Summer is the best time of the year for me, because I like the heat a lot and being able to go to the beach").

对我来说夏天是一年里最好的时候，因为我特别喜欢在温暖的天气里去海滩。

It is my favorite s_____.

Meaning recognition knowledge of the form–meaning link

This task followed a multiple-choice format were participants had to select the correct meaning from four options. Distractors were single words with the same PoS and a frequency of ≤ 2,000 word families (Nation & Webb, 2011). An 'I don't know' option was included in order to minimise guessing (Zhang, 2013).

It is the best *season*.

a) Animal
b) Time
c) Appearance
d) Place
e) I don't know

Form recall knowledge of derivatives

In this test, participants had to write down the derivative forms of the target word that were appropriate in four sentences written to constrain word class. Participants were reminded that the form of a word sometimes does not change for different word classes, and that some words may not exist in all the word classes, in which case they were instructed to indicate it with an *X*.

Season

Noun	In this country, each _____ is clearly different.
Verb	In this country, the temperature variations _____ clearly.
Adjective	In this country, the _____ temperature variations are clearly different.
Adverb	In this country, the temperature variations occur _____.

Form recognition knowledge of derivatives

This measure consists of a multiple-choice task with multiple answers. The learners were given eight different derivative options for each target word, with one correct option for each word class (or *X* if appropriate).

Season

a. *Season*	***b.*** *Seasonize*	***c.*** *Seasonally*	***d.*** *Seasonation*
e. *Seasonate*	***f.*** *Seasonal*	***g.*** *Seasony*	***h.*** *X*

Noun	In this country, each _____ is clearly different.
Verb	In this country, the temperature variations _____ clearly.
Adjective	In this country, the _____ temperature variations are clearly different.
Adverb	In this country, the temperature variations occur _____.

Meaning recall knowledge of multiple meanings

In this test, learners were assessed on their knowledge of three meaning senses of each target word through a written open-question task. For each item, they were presented with the target word, plus the word class, and a hint about each of the three meanings tested. After the hint, they were given a space to write, in their L1 or L2, a description, a translation, a definition, a synonym, or a sentence in which the specific meaning tested was used clearly, as determined by the main rater. Two additional raters assessed a random 20% of the tests (inter-rater agreement via intra-class correlation was 96%).

Season

(Noun= year)_____

(Verb= cooking)_____

(Noun= animals in season)_____

Meaning recognition knowledge of multiple meanings

In this test, the target word appeared in five sentences, with a different meaning in each. Three of those sentences represented the three meanings tested in the recall test and in the other two sentences, the word was used with an invented meaning, acting as distractors. Participants were instructed to choose all the sentences in which the word was being used with a correct meaning.

Season

a) The four seasons are winter, spring, summer and autumn.
b) The car´s season breaks very often.
c) Their dog is in season and can´t go out.
d) It is important to check the season of your computer once a year.
e) I forgot to season the fish with salt and pepper.
f) I don´t know

Form recall knowledge of collocates

In this task, participants were given a short context in their L1 and had to fill-in the blank in the English sentence with the appropriate collocate (*MI* ≥ 3, word span ±4 in COCA) of the underlined target word given the first letter (the example below translates as "Peak season is when most people go on holiday").

最多人选择出游的时期。

When you plan to go on holidays you should bear in mind that hotels are always more expensive in p_____ *season*.

Form recognition knowledge of collocates

This test took the format of a multiple-choice task, where learners were presented with a sentence in which the target word was underlined and they had to select the appropriate collocate from four options. The distractors were non-collocates of the words (*MI* non-existent or ≤ 1) as retrieved from the whole COCA (Davies, 2008).

There are always more tourists in _____ *season*.

a) Main
b) Peak
c) Big
d) Top
e) I don't know

Procedure

After receiving written consent from each participant, the test battery was administered in a pen-and-paper format to small groups of participants or individually, depending on their availability. Each test began with the instructions for completion and an example illustrating how to respond to the items. All tasks were administered individually following the order of administration piloted and used by González-Fernández and Schmitt (2020): form-meaning link form recall → VLT 5K/3K → form-meaning link meaning recognition → derivatives form recall → derivatives form recognition → multiple-meanings recall → collocate form recall → VLT 10K/2K → multiple-meanings recognition → collocate form recognition. Finally, participants handed in each separate task to the author before starting the next one.

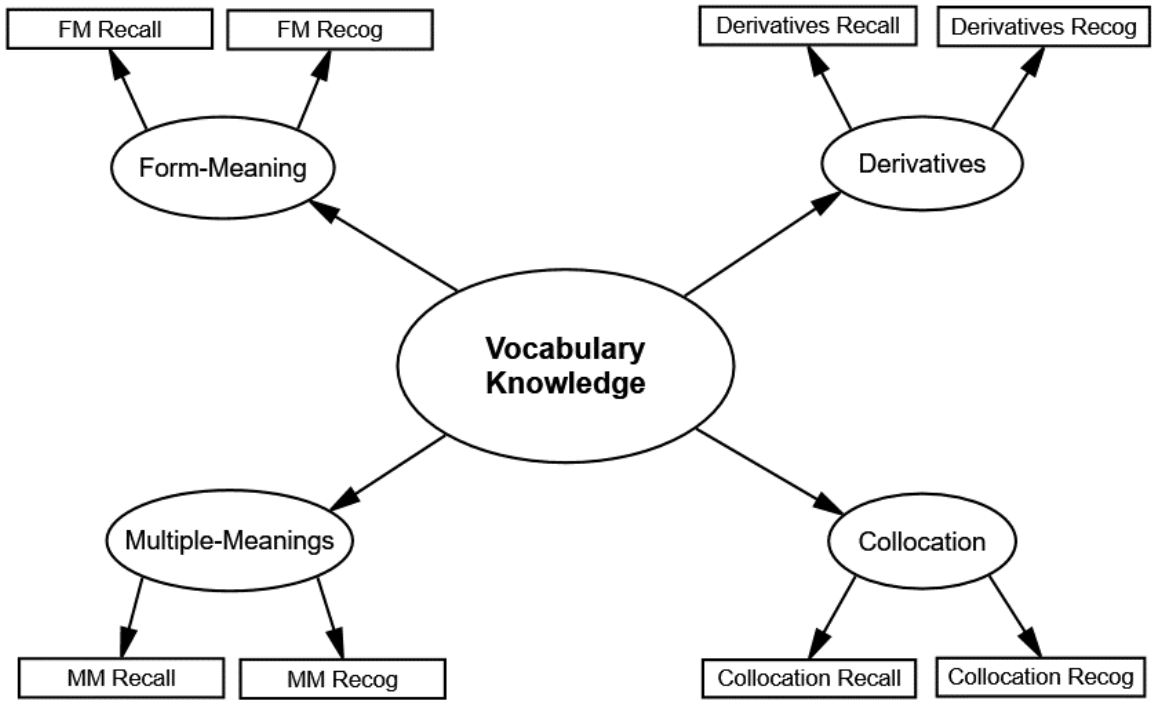
Analyses

Given its theory-driven and validating nature (Kline, 2016), confirmatory factor analysis (CFA) was employed to provide validity evidence for two opposing models about the dimensionality of vocabulary knowledge: a multidimensional and a unidimensional model.

The data analysis involved three main stages. Firstly, second-order[4] CFA was used to examine the plausibility of the multidimensional nature of vocabulary knowledge in a group of Chinese-speaking EFL learners by testing the model in Figure 2. This hypothesised multidimensional model follows the dimensional approach, specifically Nation's framework. According to this framework, the general vocabulary-knowledge construct encompasses

different word-knowledge dimensions which, in turn, are composed of receptive and productive levels of mastery. To illustrate this description, in this model the four word-knowledge components are hypothesised to reflect independent dimensions, each represented by their recall and recognition levels of mastery, which load into the second-order construct of vocabulary knowledge[5]. The observed measures (i.e., recall and recognition indicators) were specified as composite scores of the results for all the items in each test.

Figure 2 Hypothesised multidimensional model of vocabulary knowledge



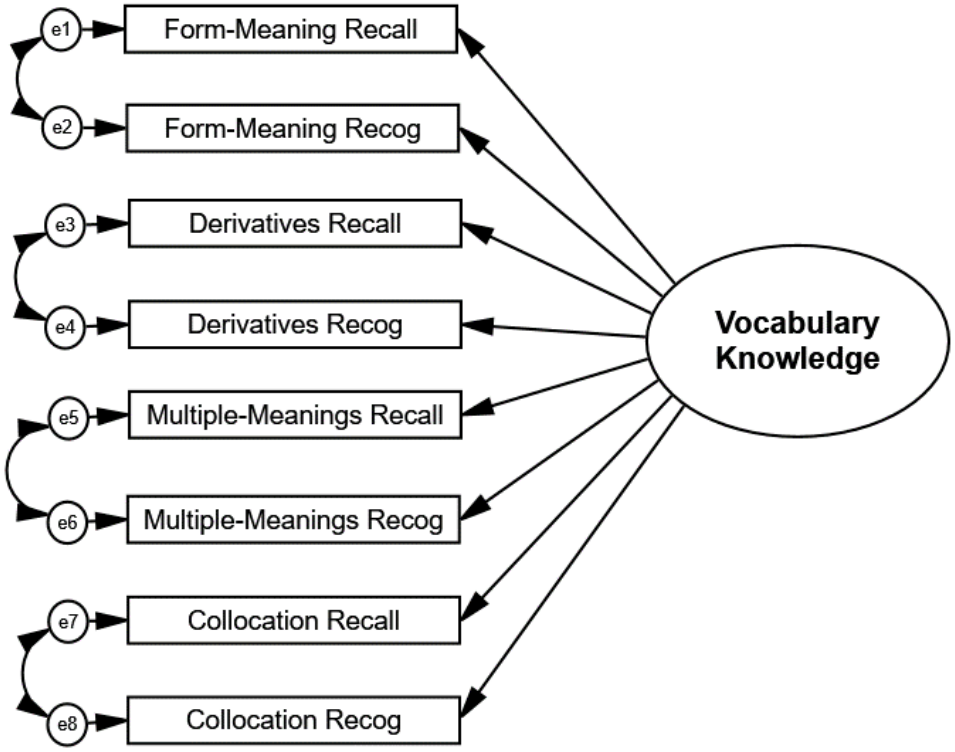
Notes: Ovals represent latent factors (i.e., unobservable variables comprised of multiple observed measures), boxes represent indicators (i.e., observed measures) and arrows represent regression paths.

Secondly, first-order CFA was employed to test whether the unidimensional description of vocabulary knowledge previously found in research (i.e., González-Fernández & Schmitt, 2020; Spencer et al., 2015) applies to the L1-Chinese group of learners. The unidimensional model for this study is illustrated in Figure 3. As can be seen, vocabulary knowledge is conceptualised as a single factor represented by each of the word-knowledge aspects (i.e., observed measures) defined as the composite scores in each task. Residual correlations are

specified between the recall and recognition tasks for each component in order to account for the intrinsic interrelationship between these aspects and measures (Schmitt, 2014).

Finally, multigroup CFA was applied to investigate whether the dimensionality of the vocabulary-knowledge construct differs as a function of L1 background, specifically the L1 cognate status, and to identify similarities and differences in the behaviour of each word-knowledge aspect across the two EFL learner populations (i.e., Chinese- and Spanish-speaking EFL learners). If a common model is found to be invariant across the samples, generalisability of the construct for the learner groups can be claimed (Kline, 2016).

Figure 3 Unidimensional model of vocabulary knowledge



The data was analysed using the package *lavaan* version 0.6-5 (Rosseel, 2012) in *R* (version 3.6.2) with the MLR estimator (data uni- and multivariate non-normal).

Results

Preliminary analyses

The descriptive statistics for the whole sample (i.e., L1-Chinese and L1-Spanish learners combined) are presented in Table 1 (see Appendix 2 in the On-line Supplementary Materials for the descriptive statistics by language group).

Table 1 Correlations, means (%), range and internal consistency reliability for the vocabulary measures across the whole sample ($N = 314$)

	FM Recall	FM Recog	Deriv Recall	Deriv Recog	MM Recall	MM Recog	Collo Recall	Collo Recog
FM Recall								
FM Recog	.638**							
Deriv Recall	.694**	.705**						
Deriv Recog	.663**	.675**	.870**					
MM Recall	.687**	.683**	.794**	.758**				
MM Recog	.601**	.733**	.722**	.685**	.772**			
Collo Recall	.626**	.649**	.703**	.675**	.713**	.683**		
Collo Recog	.575**	.642**	.680**	.626**	.678**	.650**	.772**	
M	53.36	80.96	51.60	63.65	50.01	67.60	58.79	78.41
SD	15.52	11.78	15.99	15.78	13.14	13.67	17.22	14.96
Range.	10-95	40-100	5-91.2	18-97.5	6.6-90	28.3-96.7	15-100	25-100
α	.72	.64	.92	.90	.87	.85	.73	.70

Spearman: $^{**}p < .01$

FM Recall = Form–meaning Recall; FM Recog = Form–meaning Recognition; Deriv Recall = Derivative Recall; Deriv Recog = Derivative Recognition; MM Recall = Multiple-Meanings Recall; MM Recog = Multiple-Meanings Recognition; Collo Recall = Collocation Recall; Collo Recog = Collocation Recognition.

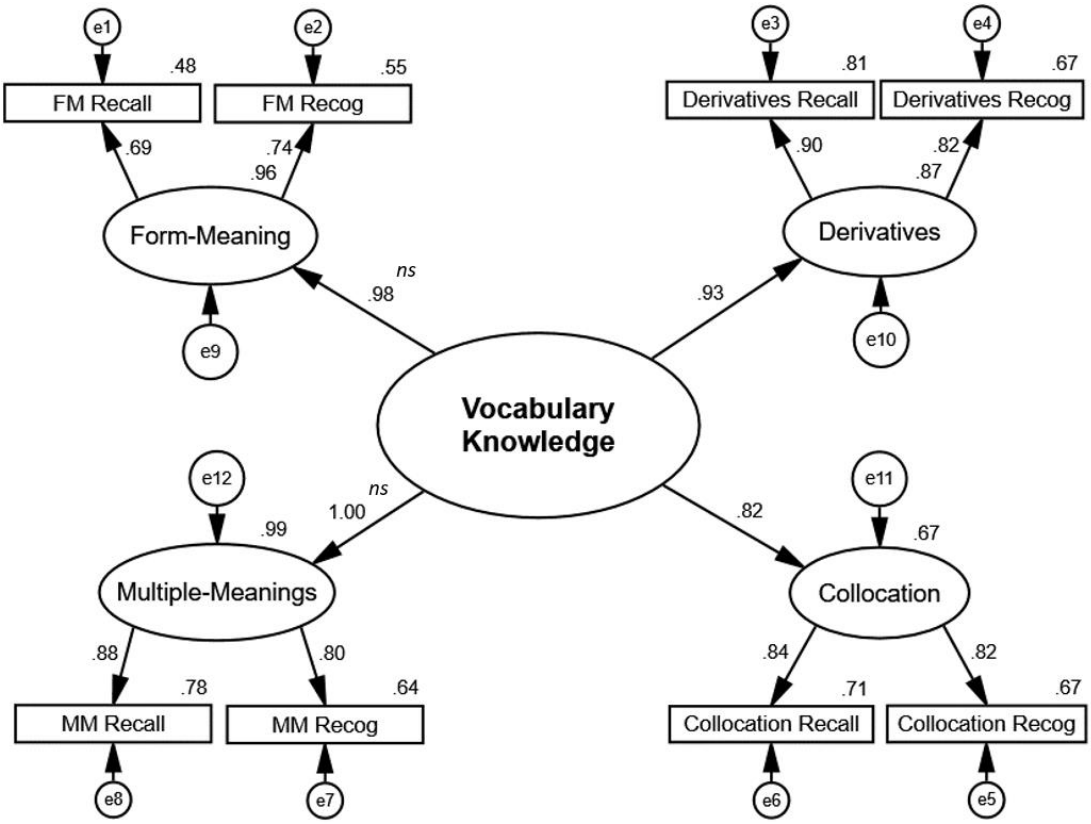
Table 1 shows that, on average, the overall sample knew at least half of the items for each test, but the scores ranged considerably across participants and some aspects were better known than others. In line with much previous research (see literature review above), correlations between the word-knowledge aspects range from .575 to .870, showing a high

degree of relationship among the vocabulary measures ($r \geq .60$, Plonsky, 2015), but lower than the threshold for concerns about multicollinearity ($r \geq .90$, Kline, 2016). Internal consistency was high (≥.70, DeVellis, 2012), for all the tasks (Cronbach's α= .70-.92), except for form-meaning recognition (α= .64)[6].

Multidimensional model of vocabulary knowledge

Prior to testing the hypothesised multidimensional model (Figure 2), preliminary analyses were conducted in order to validate the word-knowledge measures for the L1-Chinese sample (see these analyses in Appendix 3 in the On-line Supplementary Materials). After ensuring that the measures and constructs were valid, the multidimensional model was fitted using second-order CFA. Figure 4 shows the strength of the relationships between the word-knowledge dimensions and vocabulary knowledge as described in this model, and Table 2 presents its model fit.

Figure 4 *Model 1*: Multidimensional model of vocabulary knowledge for Chinese EFL learners



Z scores and significance levels for each path in Appendix 3. All paths present standardised regression coefficients.

Table 2 Fit indices for the multidimensional (*Model 1*) and unidimensional (*Model 2*) models

	χ^2	*df*	*p* value	χ^2/df	CFI	RMSEA (90% CI)	SRMR	AIC	BIC	Adjusted BIC
Acceptable fit	*The smaller, the better*		*>.05*	*1-3*	*>.95*	*<.05 / <.08*	*<.08*	*The smaller, the better*		
Model 1 fit	16.117	16	.45	1.01	1.00	.01 (.000-.072) *p* = .80	.021	6994.5	7082.3	6993.6
Model 2 fit	15.476	16	.49	0.97	1.00	.00 (.000-.068) *p* = .84	.019	6978.5	7041.2	6977.9

N =170

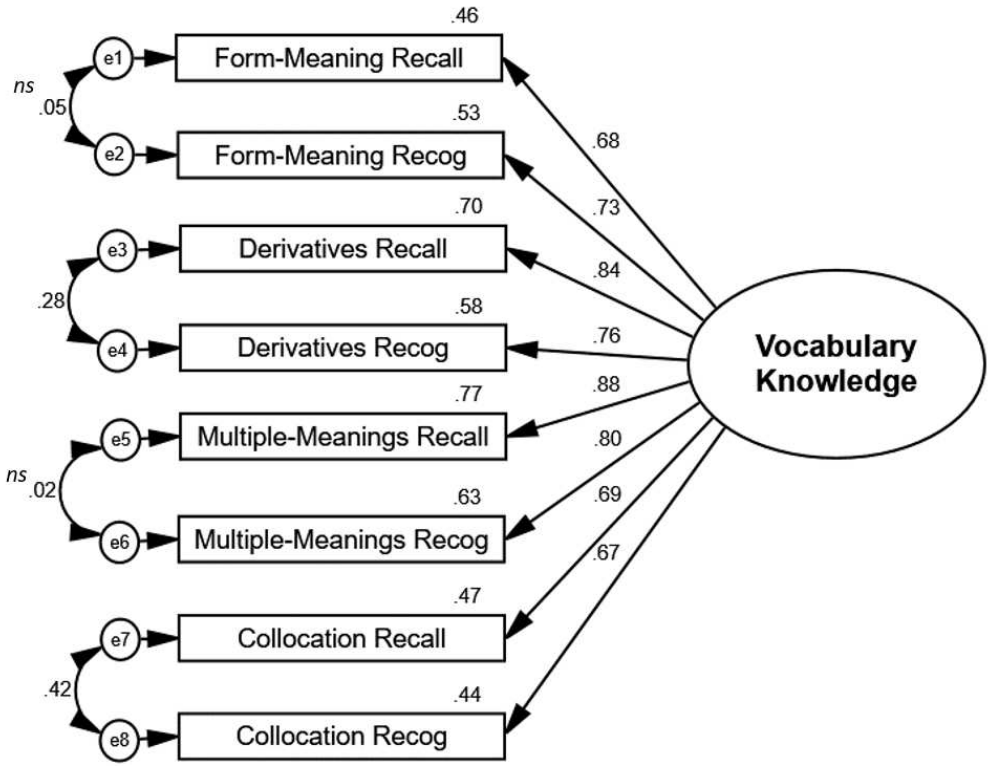
According to these global fit indices, Model 1 shows a good overall fit with the data, exhibiting construct validity. However, an examination of the relationships between the latent variables in the model showed very high (≥.85) regression coefficients among the hypothesised dimensions (some as high as β = .98 or 1.00). This indicates that there is a lack of discriminant

validity of the latent word-knowledge components (Kline, 2016), which also affected the significance of some paths (i.e., form-meaning and multiple meanings). The nonsignificant coefficients further suggest that these two latent components cannot be distinguished from the higher-order factor: vocabulary knowledge. In other words, the hypothesised word-knowledge dimensions in the model seem to be so highly interrelated that they cannot be empirically separated from each other; thus, they are redundant as independent latent dimensions and should be removed from the model (Awang, 2012). This result indicates that the multidimensional model of vocabulary knowledge is not supported by the data from the L1-Chinese learners.

Unidimensional model of vocabulary knowledge

The unidimensional model of vocabulary knowledge illustrated in Figure 3 was then tested using CFA. This model eliminates the redundant latent dimensions, and represents vocabulary knowledge as a single construct reflected by the individual recall and recognition aspects for each word-knowledge component. Figure 5 and Table 2 show the results of the evaluation for this unidimensional model for the L1-Chinese group.

Figure 5 *Model 2*: Unidimensional model of vocabulary knowledge for Chinese EFL learners



Z scores in Appendix 3. All values $p<.001$, except *ns*=nonsignificant ($p=.71$ and .88). Paths present standardised coefficients.

As shown in Table 2, Model 2 fitted the data well, demonstrating construct validity. It also shows slightly better fit indices than the multidimensional model (i.e., smaller χ^2, χ^2/df, RMSEA, SRMR, AIC and BIC), providing initial support for the unidimensional model in detriment of the multidimensional one. Figure 5 shows that all factor loadings were positive and ≥.40, considered as evidence of unidimensionality (Hooper et al., 2008). All loadings were statistically significant and the average variance explained for the model (AVE) exceeded .50 (AVE =.58), showing that convergent validity was also achieved by this model (Awang 2012). Finally, composite reliability showed very good construct reliability for the model (CR =.92).

The residual correlations between the recognition and recall aspects were not significant for the form-meaning and multiple-meanings components, but they were for derivatives and collocation. This suggests that the recall and recognition aspects for the latter two components are measuring some common knowledge not included in the model. It is unlikely that this

shared knowledge is purely word-based, otherwise the model would have captured it. Thus, a logical explanation is that it is representing certain extra-lexical linguistic information that influences vocabulary knowledge and was tapped into by the derivatives and collocation measures, such as meta-morphosyntactic information for using derivatives in context and intuitions of natural use for collocations.

This analysis demonstrates that the unidimensional model of vocabulary knowledge is valid and reliable and, unlike the multidimensional model, presents an appropriate internal structure. Therefore, its suitability as an empirical representation of vocabulary knowledge is supported for these EFL learners.

In conclusion, the unidimensional model has been found to be the best representation of vocabulary knowledge for the current sample. This indicates that vocabulary knowledge might be better seen as a unidimensional construct for L1-Chinese EFL learners, where all the word-knowledge aspects are essential in explaining this construct and contribute highly to overall vocabulary knowledge, but they do not function as independent dimensions. This unidimensionality means that the various types of word knowledge do not behave as autonomous and entirely detachable entities.

This result corroborates the finding by González-Fernández and Schmitt (2020) that a unidimensional vocabulary-knowledge model provides the best representation of the construct for a different EFL learner group (L1 Spanish). This shared outcome suggests that there seems to be a common trend in the behaviour of vocabulary knowledge across the two groups independently, implying that a unidimensional description might be a better empirical representation of L2 vocabulary knowledge than a multidimensional interpretation for EFL learners in general. However, in order to explore the generalisability of this model, that is, whether the unidimensional construct is empirically invariant and the word-knowledge aspects function in a similar manner across different groups of EFL learners, a simultaneous statistical

comparison of the best-fitting model across both groups is needed (Steinmetz et al., 2009). Multigroup CFA provides the means for this more comprehensive and reliable simultaneous comparison of the best-fitting, unidimensional model across the two dissimilar L2 language groups, L1-Spanish and L1-Chinese EFL learners.

Multigroup CFA of the unidimensional model

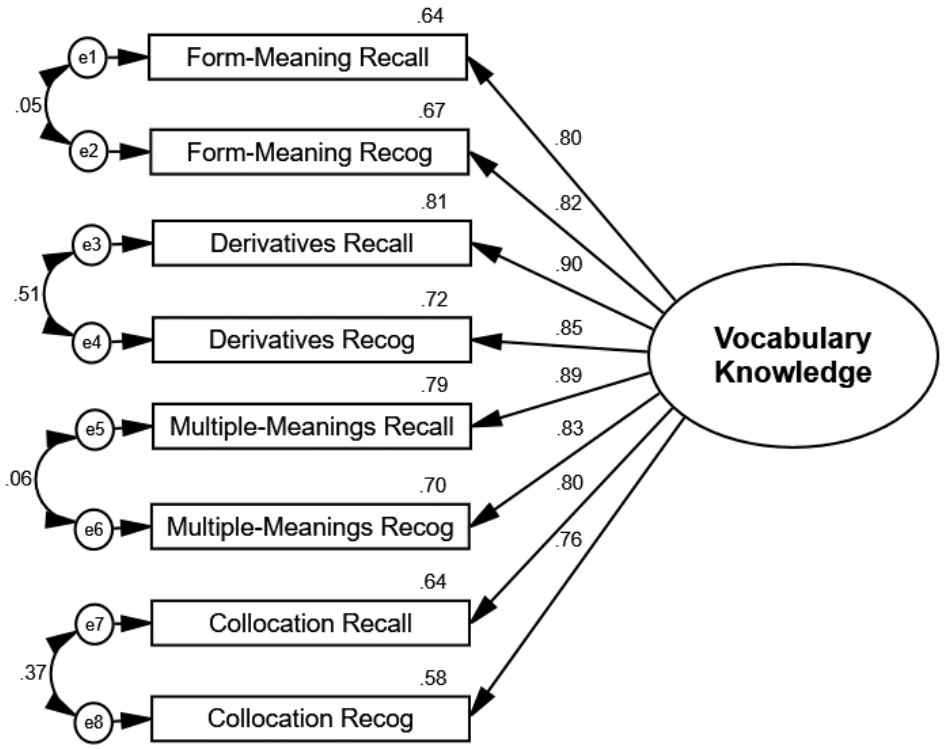
Multigroup CFA was employed to empirically compare the behaviour of the unidimensional model (see Figure 3) across the L1-Chinese (n = 170) and L1-Spanish (n = 144) EFL learners concurrently. This multigroup comparison allows us to establish whether the unidimensional vocabulary-knowledge model is group-invariant, that is, functions equivalently across the two groups. Examining this model invariance involves comparing a sequence of increasingly restrictive hierarchical models with parameters fixed to be equal across groups to an analogous model where parameters are freely estimated for each group, and testing whether the fixed estimation of parameters affects the model fit (Steinmetz et al., 2009). This analysis assesses model invariance at the level of factor loadings (weak invariance), intercepts (strong invariance) and residual variances (strict invariance). Following previous research guidance (Byrne, 2016; Cheung & Rensvold, 2002), the overall model fit and the CFI ratio test (ΔCFI)[7] were employed to check whether each subsequent model significantly affected the goodness-of-fit of the previous model. A $\Delta CFI \leq -.01$ illustrates a similar goodness-of-fit, and thus is considered as evidence of invariance between the models for the two groups (see Appendix 4 in the On-line Supplementary Materials for a detailed description of this analysis).

The results of the examination of model invariance across groups confirmed partial strict measurement invariance of the vocabulary-knowledge construct and its word-knowledge aspects for both the Chinese and the Spanish populations ($\Delta CFI = -.008$). This implies that, in general, the word-knowledge aspects as specified in this study measure the same construct,

have the same meaning and are able to gauge overall vocabulary knowledge with a similar degree of precision in each language group (Kline, 2016). That is, the unidimensional vocabulary-knowledge model was found to be generally invariant across EFL learner groups from cognate and non-cognate languages.

Given evidence of measurement invariance, Figure 6 presents the results of the multigroup evaluation of the best-fitting, unidimensional model across the whole sample (Appendix 4 shows the results per language group). Consistent with the single-group findings, multigroup CFA showed that the unidimensional model fitted the data well for the general EFL learner sample (χ^2 = 41.87, *df* = 32, *p* = .114, CFI = .995, RMSEA = .04, 90%CI [.000-.078], SRMR =.016). This result demonstrates the validity of this model as a representation of vocabulary knowledge across EFL learners simultaneously. As in the previous unidimensional model, all the factor loadings were significant and loaded highly (>.60, Awang 2012) on the vocabulary-knowledge construct (*R* between .76 and .90), explaining between 58% and 81% of the variance in overall vocabulary knowledge.

Figure 6 Multigroup unidimensional model of vocabulary knowledge for Spanish and Chinese EFL learners



N = 314. All paths present standardised regression coefficients.

Overall, the results of the multigroup CFA show that the unidimensional model does not differ as a function of the cognate status of learners' L1. This multigroup comparison provides evidence for the unidimensional structure of vocabulary knowledge to be equivalent across both language populations, such that the two groups behave similarly with respect to the general vocabulary-knowledge construct and its various word-knowledge aspects despite their L1-related differences.

Discussion

The current study explored the structure of relationships among L2 vocabulary-knowledge aspects across EFL learners from diverse L1 backgrounds, yielding the following three main findings with implications for understanding, teaching and testing vocabulary knowledge.

Empirical unidimensionality of vocabulary knowledge

First, the study investigated whether vocabulary knowledge behaves consistently with the hypothesised multidimensional descriptions of the construct in an unexplored group of L1-Chinese EFL learners. It was found that the word-knowledge components that have been theoretically conceptualised as distinct dimensions of vocabulary knowledge were not empirically independent from each other, suggesting that the various word-knowledge aspects do not exist in isolation. This result is in line with previous research which also found no empirical evidence for the multidimensionality of vocabulary knowledge in another L2-adult group of English learners as well as in a sample of L1-English children (González-Fernández & Schmitt, 2020; Spencer et al., 2015, respectively). Together, these findings indicate that, despite the descriptive convenience of the multidimensionality of vocabulary knowledge, the various aspects of word knowledge may not behave as independent dimensions in either L1 or L2 English, although further research is needed to corroborate these results.

Instead, the current study found support for a unidimensional model which conceptualises vocabulary knowledge as a single construct comprised of the recall and recognition types of word-knowledge. This result corroborates and provides further evidence for the unidimensionality of vocabulary knowledge previously found for L1 English (Spencer et al., 2015) and Spanish EFL learners (González-Fernández & Schmitt, 2020), and extends its validity as the best-fitting description of the construct to L1-Chinese EFL learners. In this model, the various word-knowledge types were found to interrelate very highly with each other (reflected in the high factor loadings), indicating that there exist links among them such that what one knows about a word influences all other types of knowledge of that word (Spencer et al., 2015). Thus, this unidimensionality and its associated high factor loadings indicate that these word-knowledge aspects cannot be considered independent from each other, but rather different levels of knowledge of the same underlying construct. This means that knowledge of

one aspect (e.g., form-meaning recognition) influences, to a great extent, knowledge of other aspects (e.g., derivative recall). This finding is consistent with the above-reviewed research exhibiting high intercorrelations among lexical components when assessing the same word-sets across tasks (e.g., Chui, 2006). In addition, employing CFA methodologies, Kremmel et al., (2017) found that an L2 phraseological-knowledge factor correlated at .90 with a general vocabulary factor while measuring different word-sets. Similarly, Koizumi and In'nami (2020) found a striking correlation of .94-.95 between an L2 form-meaning link factor and a depth factor comprised of polysemy, collocations and associations. The unidimensional model revealed in the current study provides evidence that the word-knowledge aspects do not behave as entirely separable entities, and explains why the correlations exhibited by much previous research are so high that they can barely distinguish the lexical aspects. Importantly, the current findings also emphasise the need to consider the model's internal structure as well as its fit to make well-substantiated interpretations of lexical dimensionality.

This unidimensionality result, however, goes counter to the two studies which reported a three-dimensional and a two-dimensional model as the preferred descriptions of L1 and L2 vocabulary knowledge (Kieffer & Lesaux, 2012, and Koizumi & In'nami, 2020, respectively). Certain distinctions between the measures employed in these studies and the present one can inform about the different results. While in the current study all the measures were purposefully designed with the aim to explore the nature of vocabulary knowledge in mind, Kieffer and Lesaux's and Koizumi and In'nami's studies utilised ready-available and independent tasks which had been developed for a different purpose. This means that they could not control for the differential features of the various word-sets included in each vocabulary measure and match them across tasks. It is possible that having assessed knowledge of distinct sets of words across vocabulary components made some tasks tap into additional factors derived from the differential characteristics of each word-set, thus generating independent dimensions and

leading to the multidimensionality finding. For example, Kieffer and Lesaux used a measure comprised of unknown words for their L2 and L1 participants (thus, low-frequency and/or specialised vocabulary), and required learners to guess them from context. This measure alone clustered into an independent dimension, which suggests that this extra dimension is likely a result of the measure itself assessing less common words than the other measures and tapping into a different linguistic ability (i.e., the reading strategy of inferring meaning from context), rather than strictly into vocabulary knowledge. Similarly, the morphological-awareness factor in their study contained a task assessing nonword derivation. It is possible that the fake nature of the items targeted in this task affected students' performance to the extent of leading to an independent dimension. In this sense, the multidimensionality results in these studies might have been an artifact of the ready-available measures and the particular features of the distinct word-sets tested across the various vocabulary tasks, which might have inadvertently tapped into additional constructs. This claim is supported by Stewart et al.'s (2012) research, who found that two misfit target words in a vocabulary task clustered together under an independent dimension distinct from the other words being tested, which in turn led to the weak multidimensionality of the task. In contrast, the studies that have found evidence for the unidimensionality of vocabulary knowledge (i.e., González-Fernández & Schmitt, 2020; Spencer et al., 2015), including the present study, designed the vocabulary measures for the purpose of assessing lexical dimensionality and controlled for the effect of the differential characteristics of the items assessed in each word-knowledge aspect, which typically involved examining the same target words across the multiple word-knowledge measures. While this approach does not allow for a total random sampling of words, it ensures the consistency of target-item features across tasks and reduces the risk of overdimensionalisation due to differential word characteristics (Stewart et al, 2012). In addition, Spencer et al. found unidimensionality of L1-English vocabulary knowledge when assessing different as well as the

same target words across purposefully-developed tasks. This suggests that the unidimensionality finding in the current study is unlikely to be simply due to the assessment of the same word-set across different measurements, and emphasises the need to use measures developed for the specific purpose of assessing word-knowledge dimensionality to prevent overdimensionalisation. It should be noted that Spencer et al.'s study is restricted in that some of their measures did not correlate as highly as it would be expected for aspects of the same construct. While this lower intercorrelation is probably due to the great variety of formats and administration modes employed across tasks, it further stresses the importance of thoroughly controlling the lexical measures utilised in dimensionality research.

Taking these results together, it becomes apparent that assessing the various word-knowledge aspects by means of ready-available tasks which measure distinct word-sets that do not control for the features of the target words affects the vocabulary dimensionality results, and increases the risk of tapping into other related constructs or skills resulting in the emergence of extra dimensions. This presents issues for construct validity and hinders the interpretability of the results. This finding has important implications for theory and measurement, as it suggests that assessing how much learners know about various aspects of the same words seems to be a more theoretically-sound method to guarantee that we are assessing only the intended construct when the aim is to explore the dimensionality of vocabulary knowledge. Consequently, following the traditional description and operationalisation of word knowledge, which is described as mastery of various types of knowledge for each word, the present study indicates that vocabulary knowledge should be considered a unidimensional construct in practice.

Towards a generalisable conceptualisation of vocabulary knowledge

Multigroup CFA found support for the validity of the unidimensional vocabulary-knowledge model across two groups of EFL learners from a cognate (Spanish) and a non-cognate (Chinese) L1 backgrounds, revealing that it is a better empirical representation of L2 vocabulary knowledge for EFL learners from very distinct L1s. Importantly, this multigroup examination also confirmed that the various word-knowledge aspects and the vocabulary-knowledge construct represent the same underlying concepts and function equivalently in both EFL learner groups. This implies that the word-knowledge aspects follow a similar pattern of relationships relative to each other and the overall vocabulary construct regardless of learners' L1 (Byrne, 2016). Thus, in practice, students at a similar level of overall vocabulary knowledge would be expected to respond to the word-knowledge aspects in a similar manner regardless of their L1.

This cross-validation and group-invariance of the unidimensional vocabulary model is even more remarkable when we consider how different the L1s of the two learner groups are from each other as well as in relation to the English language. While Spanish is a close cognate language of English, sharing around 34-37% of cognate words (Lubliner & Hiebert, 2011), Chinese and English are non-cognate languages. There is substantial evidence that speakers whose L1 shares cognates with the target L2 would experience a cognate facilitative effect that might improve their overall L2 vocabulary knowledge when compared to L1 speakers from non-cognate languages (Chen et al., 2012; Vidal, 2011). Since these two EFL groups with significantly different L1s showed the same vocabulary-knowledge structure, meaning and behaviour of the word-knowledge aspects, this provides robust evidence for the generalisability of the unidimensional conceptualisation of vocabulary knowledge across EFL learners in general.

Refining Nation's word-knowledge framework

The present study can shed light on how well Nation's (2013) descriptive word-knowledge framework represents the way L2 learners structure their vocabulary knowledge in practice, allowing us to interpret and refine this theoretical framework based on empirical evidence.

Nation's description of overall vocabulary knowledge emphasises that lexical development requires achieving mastery of several types of knowledge for each word. The unidimensional model in this study validates this theorisation by showing that the four word-knowledge components tested in their recall and recognition masteries contribute highly and similarly to the general vocabulary-knowledge construct, without any one aspect being significantly lower than the others (all factor loadings are between .76 and .90 for all the EFL learners). This finding further enhances Nation's framework by empirically confirming that each individual word-knowledge aspect is important and complements each other in representing overall vocabulary knowledge in a second language. This does not imply, however, that all aspects of word knowledge are equally relevant to all learners at all developmental stages (Nation, 2020). For example, in initial learning stages the form-meaning link and written/spoken form would be the most important aspects, and, as proficiency develops, aspects such as morphological or collocational knowledge would gain more relevance. Thus, teachers should focus on the aspects of word knowledge that are most useful for learners to develop in order to achieve their specific learning goals. Importantly, the present study also substantiates that the most appropriate method to examine the dimensionality of vocabulary knowledge while only assessing the intended construct is by measuring different types of knowledge of the same set of words. This highlights the significance of Nation's description of word knowledge as a framework to research the nature of vocabulary knowledge.

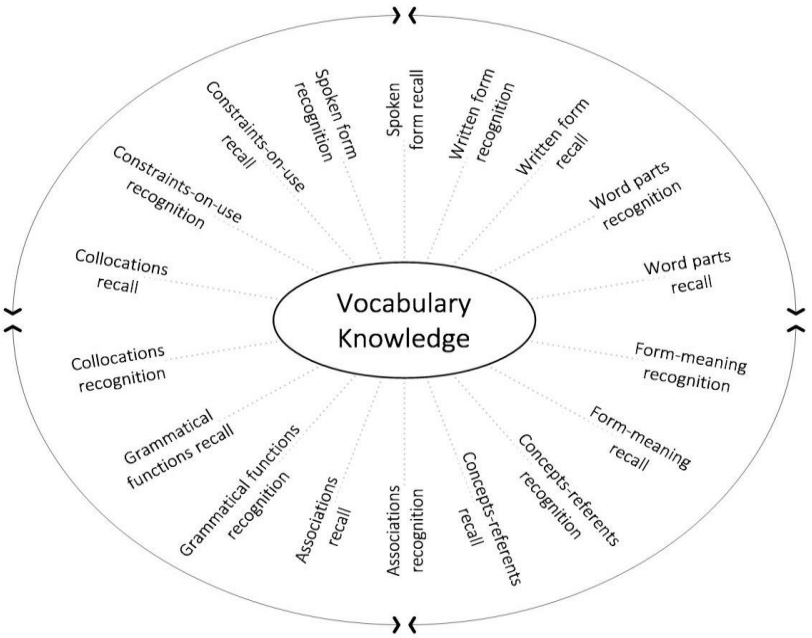
The current study also suggests that some parts of Nation's framework require further refinement. Given that the various word-knowledge aspects identified in this framework are presented separate from each other in a list or table, researchers have typically interpreted them as separable and independent (if related) dimensions, as stated by Gyllstad (2013). The findings in the present study show that this is not the case, and that the word-knowledge aspects are not known by L2 learners as distinct and isolated dimensions. Rather, they behave as intertwined types of lexical ability, where knowledge of one aspect influences knowledge of the other aspects. It also shows that EFL learners from different L1 backgrounds accumulate and structure this knowledge in parallel. Thus, although we employ labels to conceptually distinguish the various types of knowledge for convenience, it seems that the boundaries we establish between them are, to some extent, artificial. Sticking too strongly to these word-knowledge aspects and components as separate entities has made us overlook their actual commonalities and the strength of their interrelationships (Milton & Fitzpatrick, 2014, pp.176-177). Consequently, moving forward, a more holistic understanding of L2 vocabulary knowledge is needed in theory and research (see the implications section).

For pedagogical purposes, however, the discrete description of word-knowledge aspects still has value as an accessible explanation of the complexity of vocabulary knowledge. This study corroborates previous claims that the various word-knowledge aspects are known by L2 learners to different degrees. For example, the results in Table 1 revealed that the derivatives and multiple-meanings recall aspects were known at a lower level than the other aspects, suggesting that they might be more complex to develop in an L2. This indicates that, while the aspects represent a single construct, they still pose distinct learning burdens to L2 learners, and thus, it is expected that some are known/learnt before others (González-Fernández & Schmitt, 2020; Spencer et al., 2015). Therefore, teaching these aspects in isolation and at different stages in the vocabulary acquisition process is still justified and adequate for

instructional purposes, although requires an understanding that knowledge of one aspect does not accumulate independently from the other aspects.

The above findings hint that an alternative presentation of the word-knowledge framework that accommodates the available empirical evidence is warranted. As a first step to address this matter, Figure 7 extends this study's findings on eight word-knowledge aspects to conjecture a more empirically-grounded explanatory illustration of Nation's framework. In this illustration, L2 vocabulary knowledge is presented as a unique construct, where the individual recall and recognition types of word knowledge are pictured as interlinked, existing in parallel and orbiting around the word knowledge construct in no specified order. This illustration is not the final picture, however, and should be updated as further research continues investigating the nature of L2 vocabulary knowledge.

Figure 7 Empirical illustration of the word-knowledge framework



Theoretical and practical implications

The above findings have implications at two key levels: theoretical and practical. Regarding the theoretical implications, the empirical support from this study for the unidimensionality of vocabulary knowledge in EFL learners suggests that previous multidimensional conceptualisations of vocabulary knowledge should be reconceptualised. The current study found robust evidence that the various types of word knowledge do not behave as independent dimensions in the process of mastering L2 lexical knowledge. From this finding it follows that we should be cautious regarding the interpretation and terminology employed when describing this construct. In particular, in the field of vocabulary studies, research has typically referred to the word-knowledge aspects as *dimensions* of vocabulary knowledge (Daller et al., 2007; Gyllstad, 2013; Milton & Fitzpatrick, 2014). Referring to each of these types of word knowledge as *dimensions* carries certain implications: if two attributes of knowledge are considered to be two separate dimensions, this implies that they are fundamentally distinct from each other and independent in nature. Since this study found that this is not the case for the word-knowledge aspects tested, keeping the term *dimensions* to refer to them is, thus, problematic and can give rise to misleading descriptions, interpretations and applications of the construct in future research. For example, if researchers continue treating these aspects as separate entities, they may misinterpret the strong relationship between two types of word knowledge as a predictive effect of one construct over a different one; but this effect might be a consequence of the nature of the word-knowledge aspects as elements of one construct. This study shows that it is inappropriate to continue referring to these levels of ability as dimensions, and proposes that they should be better referred to as aspects or types of word knowledge.

The main practical implications of the present study concern vocabulary measurement. The findings provide an initial empirical model of the general structure of vocabulary knowledge which can inform the development and validation of vocabulary tests, and allow

for a meaningful interpretation of test scores (Koizumi & In'nami, 2020; Schmitt et al., 2020). It is important to emphasise that the unidimensional model does not suggest that assessing only one aspect of vocabulary knowledge is sufficient when the aim is to fully gauge learners' word knowledge. The findings show that each of the word-knowledge aspects is an essential part of overall vocabulary knowledge, and thus basing conclusions on the mastery of just one of these aspects provides only partial insights about lexical knowledge, and can lead to the underrepresentation of this construct. Thus, practitioners and researchers seeking to obtain a more complete and detailed understanding of L2 learners' overall vocabulary knowledge or to examine the development of different components should employ several vocabulary measures which assess knowledge of words across various aspects (Nation & Webb, 2011), such as in the battery employed in this study. The decision of which and how many aspects to measure should depend, however, on what is meaningful for the ultimate goal of teaching/research. If the aim is only to obtain a rough estimation of learners' vocabulary knowledge, for example in order to investigate its relationship with other linguistic domains, such as reading or grammar, the unidimensionality of vocabulary knowledge suggests that multiple aspects of word knowledge might not need to be measured in this case. For this purpose, researchers might choose to assess only few specific types of word-knowledge as a proxy of the general underlying construct of vocabulary knowledge. It should be acknowledged, however, that this approach might not represent overall vocabulary knowledge precisely, and that the more aspects we test, the more accurate this representation of word knowledge will be.

Overall, the present paper supports the pedagogical value of the discrete description of each word-knowledge aspect in teaching, as well as the design of separate tests to assess them when examining the development and knowledge of each aspect. But the paper also points to the need of recognising the holistic and unidimensional nature of vocabulary knowledge as an

empirically-based conceptualisation of the construct, using it to inform our interpretation of vocabulary research and test results.

Conclusion

In sum, this study has expanded previous research by examining the nature of vocabulary knowledge with purposefully-designed measures that control for differential word features, and by exploring the validity of two opposing vocabulary models across two different populations of EFL learners. However, it has some limitations that serve as foundations for interesting questions to be addressed in future research. First, while the study innovatively demonstrates the empirical generalisability of a unidimensional model of vocabulary knowledge across two groups of EFL learners from different L1s, further research is warranted to explore the vocabulary dimensionality across other EFL populations, as well as in second languages other than English. This research could investigate whether a universal model of L2 vocabulary knowledge can be drawn, which would provide a robust basis for the development and validity of a much-needed general theory of vocabulary acquisition (see Schmitt's (2019) research agenda). In addition, the current study shows that the dimensionality results might be affected by how vocabulary knowledge is conceptualised and operationalised, in particular whether or not researchers design measures purposefully and control for the effect of differential word features by assessing the same or different sets of words across tasks. However, it is unclear from the available research which specific word characteristics might interfere with vocabulary dimensionality, particularly when assessing different word-sets. Future replication research is needed to examine whether features such as the length, imageability or frequency of different target words might influence the unidimensionality of word knowledge. For example, the current study selected most target items from higher-frequency bands, so replication research could employ lower-frequency words to compare

differences and similarities in the dimensionality results. In addition, given the impracticality of measuring *all* word-knowledge aspects concurrently, future studies should investigate different combinations of aspects than the ones selected in the current paper to build a composite picture of overall vocabulary knowledge. Finally, it is conceivable that assessing the word-knowledge aspects by means of not only offline but also online measures may influence the dimensionality of vocabulary knowledge (see Godfroid, 2020 for a review on how to implement this approach). Additional research is needed to corroborate this claim by examining the dimensionality of vocabulary knowledge while employing online and offline lexical measures. This research would paint a fuller picture of the overall nature of vocabulary knowledge.

Notes

1. The terms *construct* and *dimension* refer to independent unobservable concepts that can only be gauged through observable measures. They are typically considered to be related hierarchically, but have been employed inconsistently in the vocabulary literature: for some authors *construct* is the high-level concept (e.g., Henriksen, 1999) and for others *dimension* is the superordinate notion (e.g., Gyllstad, 2013). Following descriptions of the components approach, in this article, the term *dimension* is employed as a subordinate of a *construct*, so that vocabulary knowledge is regarded as a construct and the vocabulary components are considered dimensions within it.
2. In this article, the term *component* refers to the word-knowledge dimensions (e.g., derivatives or collocation), and *aspect* and *type* to the separate recognition/recall levels of knowledge (e.g., derivative recall, derivative recognition).
3. Vocabulary depth is understood in this study as learners' knowledge of multiple aspects of knowledge of individual words beyond the form-meaning link (Schmitt, 2014).
4. First-order models are comprised of first-order factors, which are represented through indicators. Second-order models include first *and* second-order factors, the latter being superordinate constructs represented by the first-order factors and their indicators (Kline, 2016).
5. This second-order model is the one examined in prior L2 lexical dimensionality research (i.e., González-Fernández & Schmitt, 2020). Yet, it is not the only possible interpretation of word knowledge. After exploring other valid representations of Nation's framework (i.e., a bi-factor model and a four-factor model), this second-order model showed an advantage in fit with lower AIC and BIC values (bi-factor model's AIC = 7002.5 and BIC = 7102.8; four-factor model's AIC = 6996.0 and BIC = 7090.1). This indicates that the second-order model is a more parsimonious interpretation of this framework, and thus, it was the model chosen to test lexical multidimensionality in this study.
6. The slightly lower α level for the form-meaning recognition task indicates more variation across participants in their responses to this test. Following Kieffer and Lesaux (2012) and Spencer et al.

(2015), it was considered appropriate to include this measure in the analysis, since CFA extracts the available common variance from lower-reliability tasks.

7. To avoid the limitations of the traditional likelihood ratio test (i.e., sensitive to sample size and very stringent), the ΔCFI has been suggested as a more appropriate goodness-of-fit index to establish measurement invariance, since it is independent from sample size and model parameters.

References

Awang, Z. (2012). *A handbook on SEM: Structural equation modelling using AMOS graphics*. Universiti Teknologi Mara.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.

Chen, X., Ramirez, G., Luo, Y. C., Geva, E., & Ku, Y.-M. (2012). Comparing vocabulary development in Spanish-and Chinese-speaking ELLs: The effects of metalinguistic and sociocultural factors. *Reading and Writing*, *25*(8), 1991–2020.

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, *35*(1), 3–25.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.

Chui, A. (2006). A study of the English vocabulary knowledge of university students in Hong Kong. *Asian Journal of English Language Teaching*, *16*, 1–23.

Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *The Journal of Educational Research*, *36*(3), 206–217.

Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. University Press Cambridge.

Davies, M. (2008). *The corpus of contemporary American English: 450 million words, 1990-present.* Retrieved from http://corpus.byu.edu/coca/.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Sage publications.

Godfroid, A. (2020). Sensitive measures of vocabulary knowledge and processing: Expanding Nation's framework. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 433–453).

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.

Gyllstad, H. (2009). Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 153–170). Palgrave Macmillan.

Gyllstad, H. (2013). Looking at L2 vocabulary knowledge dimensions from an assessment perspective–challenges and potential solutions. In C. Bardel, C. Lindquist, & B. Laufer (Eds.), *L2 Vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis.* (Vol. 2, pp. 11–28). Eurosla Monograph Series.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, *21*(02), 303–317.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing*, *25*(2), 347–373.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.

Koizumi, R., & In'nami, Y. (2020). Structural Equation Modeling of Vocabulary Size and Depth Using Conventional and Bayesian Methods. *Frontiers in Psychology*, *11*, 618. https://doi.org/10.3389/fpsyg.2020.00618

Kremmel, B., Brunfaut, T., & Alderson, J. C. (2017). Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics*, *38*(6), 848–870.

Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics*, *36*(5), 611–634.

Lubliner, S., & Hiebert, E. H. (2011). An Analysis of English–Spanish Cognates as a Source of General Academic Language. *Bilingual Research Journal*, *34*(1), 76–93.

Milton, J., & Fitzpatrick, T. (2014). *Dimensions of vocabulary knowledge*. Palgrave Macmillan.

Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, *28*(2), 291–304.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nation, P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 15–29). Routledge.

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*(3), 298–320.

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113–138.

Plonsky, L. (2015). Statistical power, p values, descriptive statistics, and effect sizes. A "back-to-basics" approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research.* (pp. 23–45). Routledge.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*(3), 513–536.

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913–951.

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

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*(1), 109–120.

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.

Spencer, M., Muse, A., Wagner, R. K., Foorman, B., Petscher, Y., Schatschneider, C., Tighe, E. L., & Bishop, M. D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing*, *28*(7), 959–988.

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity*, *43*(4), 599–616.

Stewart, J., Batty, A. O., & Bovee, N. (2012). Comparing multidimensional and continuum models of vocabulary acquisition: An empirical examination of the vocabulary knowledge scale. *TESOL Quarterly*, *46*(4), 695–721.

Vidal, K. (2011). A Comparison of the Effects of Reading and Listening on Incidental Vocabulary Acquisition. *Language Learning*, *61*(1), 219–258.

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.

Wesche, M., & Paribakht, T. S. (1996). Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth. *Canadian Modern Language Review*, *53*(1), 13–40.

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *Tesol Quarterly*, *44*(4), 647–668.

Yanagisawa, A., & Webb, S. (2020). Measuring depth of vocabulary knowledge. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 371–386). Routledge.

Zhang, X. (2013). The I don't know option in the Vocabulary Size Test. *TESOL Quarterly*, *47*(4), 790–811.