Article:
González-Fernández, B. (2025) How is vocabulary learnt? An acquisitional sequence of L2 word knowledge. TESOL Quarterly, 59 (2). pp. 755-784. ISSN: 0039-8322

https://doi.org/10.1002/tesq.3342

tesol QUARTERLY
tesol international association

How Is Vocabulary Learnt? An Acquisitional Sequence of L2 Word Knowledge

BEATRIZ GONZÁLEZ-FERNÁNDEZ
School of English, The University of Sheffield
Sheffield, UK

Abstract

Second language acquisition (SLA) researchers have long searched for patterning in the development of linguistic elements (e.g., grammar and morphology). However, little attention has been given to the examination of systematicity in vocabulary acquisition, limiting our understanding about how overall vocabulary is learnt. The current study investigates L2 learners' lexical knowledge to explore whether there exists a consistent sequence in the acquisition of vocabulary components in second languages. Written form/meaning-recall and written form/meaning-recognition knowledge of four word-knowledge components (form–meaning, collocations, multiple-meanings, and derivatives) was assessed on 314 EFL learners from two distinct L1 backgrounds (Chinese and Spanish). Implicational and Mokken scaling analyses revealed a reliable hierarchy of knowledge of vocabulary aspects where recognition knowledge preceded recall knowledge across all components. The scale remained constant across the learner groups when explored independently (using an 80% accuracy threshold) and collectively (under both 75% and 80% accuracy thresholds). This finding indicates that the various aspects of word knowledge seem to be learnt incrementally in a consistent order by EFL learners, regardless of their L1. The study offers an empirically supported framework of word-knowledge acquisition that improves our current understanding of L2 lexical development and can serve as guidance to further systematize vocabulary instruction in the EFL classroom.

doi: 10.1002/tesq.3342

INTRODUCTION

The examination of developmental orders in the acquisition of linguistic features has been of historic interest in certain areas of second language acquisition (SLA), and researchers have discovered increasing amounts of systematicity in the ways languages are learnt. For example, research has now established that some grammatical rules (e.g., negation) are acquired in a particular developmental order by L1 and L2 learners (Ellis, 2008). Yet, this systematicity has received little attention in vocabulary studies, mainly due to the inherent complexity of word knowledge. Knowing a word involves mastery of a variety of aspects, such as its pronunciation and spelling, its meanings, and how it is used in context (Nation, 2020). While there is certain idiosyncrasy in the acquisition of particular individual words, it is possible that the various aspects involved in vocabulary knowledge follow a consistent sequence of acquisition, with some aspects being learnt before others (Schmitt, 2019).

Nation, 2022 (and his earlier 2013 edition), proposed the most exhaustive description of the range of types of explicit knowledge learners can have about words. He lists nine word-knowledge components which can be known receptively ([R] understanding words in reading and listening) and productively ([P] using words in speaking and writing; see Figure 1).

Despite the framework's influence in advancing lexical research, there is still a lack of clarity about how these word-knowledge components relate to each other (e.g., does knowledge of more meanings lead to better collocation knowledge?) and how L2 learners acquire

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?

FIGURE 1. Nation's (2022, p. 54) framework of word-knowledge components.

them (e.g., are a word's derivative forms typically learned before its collocations?). This is because research has not typically examined the knowledge and relationships between the various word-knowledge aspects in a concurrent and systematic manner (Nation, 2020; Schmitt & Schmitt, 2020).

One exception is González-Fernández and Schmitt's (2020) study, which investigates multiple word-knowledge aspects[1] in a group of L1-Spanish English-as-a-foreign-language (EFL) learners. The authors found that some lexical aspects were consistently more advanced than others and encompassed the less advanced aspects. This finding suggests that the various word-knowledge aspects might be acquired by EFL learners in a specific ordering. However, their results are preliminary and restricted to only one learner group from a cognate language.[2] Given the influence that the cognate status of learners' L1 exerts on vocabulary learning (e.g., Chen, Ramirez, Luo, Geva, & Ku, 2012), it is possible that the hierarchy of knowledge found by González-Fernández and Schmitt for L1-Spanish EFL learners does not replicate in EFL learners from other linguistic backgrounds. Consequently, the questions remain how the word-knowledge aspects are acquired by EFL learners of different language backgrounds and whether a hierarchy of knowledge of these aspects exists that remains constant across distinct language groups.

The current study aimed at addressing these gaps by examining the extent of knowledge of eight vocabulary aspects by two groups of EFL learners (one cognate and one non-cognate) and explored whether an order of acquisition of written word-knowledge aspects[3] can be established for each language population individually and combined. This investigation helps explain how vocabulary learning occurs, allowing us to better systematize when and how to introduce the various aspects to students. If a consistent order is established for each learner group but this order differs across both learner populations, this would suggest that the differences between learners' L1s affect the pattern in which the L2-English vocabulary components are acquired; thus, vocabulary instruction should be adapted for each language population. If, on the other hand, the same order is found for both groups,

[1] In this article, the term *component* refers to the word-knowledge types (e.g., derivatives or collocation) and *aspect* to the separate recognition/recall levels of mastery (e.g., derivative recall and derivative recognition).

[2] *Cognate language* is used to denote languages that share a significant number of cognate words (i.e., words that derive from the same original language). English and Spanish are considered cognate languages because 34–37% of English words are cognates with Spanish (Lubliner & Hiebert, 2011).

[3] In this manuscript, all references to word-knowledge aspects concern the written modality, so that, for example, meaning-recognition refers to written-receptive meaning recognition, not spoken-receptive meaning recognition.

this would be indicative of a consistent and generalizable sequence of acquisition of vocabulary components across EFL learners that would help facilitate the methodical learning and teaching of vocabulary in the L2 classroom (Schmitt, 2019).

L2 VOCABULARY KNOWLEDGE

The first detailed examination of the knowledge and combined interrelationships among the various word-knowledge aspects in L2s was Schmitt (1998). He explored EFL learners' development of four word-knowledge components in their productive mastery: written form, word parts (derivatives), associations, and concepts and referents (multiple meanings). He found that all components were interrelated and developed in a parallel manner, although at different paces. However, Schmitt did not find any systematic pattern of knowledge which showed that some components were consistently better known by learners than others.

Other multicomponent studies found that overall recognition knowledge tends to be higher than recall knowledge for EFL learners. *Recognition* knowledge involves being able to recognize and select correct lexical information from a number of options, whereas *recall* entails recalling word-knowledge information from memory (without options) after some kind of stimulus is given (Schmitt & Schmitt, 2020). Pellicer-Sánchez and Schmitt (2010) measured learners' knowledge of four written recall and recognition word-knowledge aspects and found that the two recognition aspects (i.e., written meaning-recognition and written form-recognition) were better known (84% and 76% of words known, respectively) than the two recall aspects (i.e., word-class recall [63% knowledge] and meaning recall [55% knowledge]). This indicates that recall knowledge (both form-recall and meaning-recall) seems to be more complex for EFL learners and thus might appear later in their lexical development. This difference between lexical recall and recognition knowledge by EFL learners has led some scholars to contend that they are distinct psychometric constructs (Stewart, Gyllstad, Nicklin, & McLean, 2024). However, these studies did not assess each component in both recall and recognition levels of mastery. Thus, they cannot inform about whether this pattern of difficulty derives from the component itself (e.g., word class) or the levels of mastery assessed (recall vs. recognition).

The first research that consistently examined two levels of mastery across components was Webb (2005, 2007). He assessed written productive (form-recognition and form-recall) and receptive (meaning-recognition and meaning-recall) knowledge of five vocabulary aspects:

written form, form–meaning mapping, syntagmatic associations (i.e., collocations), grammatical functions (i.e., word class), and paradigmatic associations. His findings show that written form tends to be the best-known aspect both receptively and productively, but the extent of knowledge for the rest of the aspects varied at each mastery level. Despite having employed the same test battery, his research did not find a consistent order of difficulty of the different vocabulary aspects which could describe vocabulary learning. The same conclusion was reached by Chen and Truscott's (2010) conceptual replication of Webb's research, although they did observe that the receptive knowledge of each component tended to be higher than their productive counterpart.

From this multicomponent research, it can be inferred that: (a) the various word-knowledge components are known to different extents by EFL learners, and (b) that there is a tendency for written-receptive and recognition mastery to be known better than written-productive and recall mastery across aspects. However, this research cannot inform about a systematic ordering in which knowledge of the vocabulary components is accumulated by EFL learners. Exploring a sequence of acquisition of vocabulary components requires specific statistical analyses (i.e., implicational/Mokken scaling) and a design that accommodates to the features of such analyses.

SEARCHING FOR ACQUISITION ORDERS OF WORD-KNOWLEDGE ASPECTS

SLA research has found robust acquisition sequences for grammatical rules (e.g., Ellis, 2008), phonemes (e.g., Trofimovich, Gatbonton, & Segalowitz, 2007), and morphemes (e.g., Andersen, 1978) by means of implicational scales. However, the complexity of examining multiple word-knowledge aspects concurrently has typically deterred researchers from investigating acquisitional orders in lexical research. Thus, the hypothesis that the various vocabulary components can also be acquired in a specific sequence, where some aspects consistently develop before others, remains largely untested.

Only a handful of studies have attempted to search for a systematic order of acquisition of vocabulary components via implicational scales. Laufer and Goldstein (2004) and Laufer, Elder, Hill, and Congdon (2004) examined the knowledge that EFL learners from various L1s had of the four aspects that comprise the form–meaning link component: passive recognition (i.e., meaning recognition), passive recall (i.e., meaning recall), active recognition (i.e., form recognition), and

active recall (i.e., form recall). They found a reliable implicational scale for this vocabulary component, suggesting the following order of difficulty, and thus expected acquisition, of the form–meaning link levels (from easier to more difficult): meaning recognition ≥ form recognition > meaning recall > form recall.

Although these studies provide valuable insights into potential acquisition orders in L2 lexical development, they are restricted to the analysis of a single vocabulary component (form–meaning link). There has been only one previous study that explored the systematicity in the acquisition of multiple word-knowledge components. González-Fernández and Schmitt (2020) assessed L1-Spanish EFL learners' knowledge of four word-knowledge components, each in recall and recognition mastery: form–meaning link, collocation, word parts (i.e., derivatives), and concepts and referents (i.e., multiple meanings). Significantly, they observed a reliable order of acquisition of these vocabulary components. In particular, form–meaning recognition was found to be the best-known aspect, whereas knowledge of multiple-meanings and derivatives were the most difficult for the learners to master, particularly in recall mastery. This finding of a lexical acquisition order is exciting, but the study was only an initial exploration of the issue and examined learners from only one L1 background (Spanish). Thus, it cannot inform about the consistency of such pattern and its transferability to other EFL learner populations. Long (1990) states that the acquisition sequences of linguistic features might experience some variation due to learners' different L1s. Lexical research has shown that the cognate status of learners' L1 (such as that of Spanish and English) can positively influence L2 vocabulary learning as compared to non-cognate L1 learners (e.g., Chen et al., 2012). Therefore, it is unclear whether EFL learners from non-cognate language backgrounds also learn the various word-knowledge aspects in a specified order. The present study aimed to compare the knowledge of various vocabulary components across two EFL learner populations from distinct L1 backgrounds (i.e., cognate and non-cognate) and to examine whether an implicational order of acquisition of these components exists for EFL learners in general.

THE STUDY

The current study investigates the extent of knowledge and sequence of acquisition of vocabulary components across EFL learners from a cognate and a non-cognate L1, both individually and collectively. It builds on the conclusions and limitations of previous multicomponent research in three main ways. First, unlike earlier multicomponent

studies (e.g., Pellicer-Sánchez & Schmitt, 2010; Schmitt, 1998), the present article examines learners' knowledge of various word-knowledge components consistently at both their recall level and recognition level of form and/, or both, meaning. This approach enables a more comprehensive account of overall word knowledge and allows us to establish any differences in knowledge by component (e.g., collocation) as well as by mastery level (i.e., recall/recognition).

Second, this study explores L2 learners' knowledge of the different lexical components not only by comparing mean score ranks (i.e., Chen & Truscott, 2010; Pellicer-Sánchez & Schmitt, 2010; Webb, 2005, 2007), but also by analyzing hierarchical relationships between the components via implicational scales. Implicational scaling considers the pattern of knowledge of components by individual learners as well as the sample as a whole (Rickford, 2002), allowing us to confidently determine whether consistent implicational relationships exist among the word-knowledge components. These implicational relationships, in turn, would indicate an expected order of acquisition of the lexical components.

Finally, given the effect of learners' L1 on their L2 vocabulary learning (Elgort, 2013), the current study tests the generalizability of the order of acquisition of lexical components proposed by González-Fernández and Schmitt (2020) to another group of EFL learners from a very distinct L1 background (i.e., Chinese). This type of validation research is scarce in the field of SLA, but it is crucial to test the robustness and reliability of findings (Marsden, Morgan-Short, Thompson, & Abugaber, 2018; Schmitt, 2019), particularly in cases when an order of acquisition of linguistic items is proposed (Long, 1990). To this aim, the study investigates whether González-Fernández and Schmitt's preliminary acquisition order of word-knowledge components holds for a new L1 learner population (Chinese), and re-analyses their L1-Spanish dataset to explore whether the findings extend to a larger EFL learner group comprised of students from two distinct L1 backgrounds. Thus, the present study further inspects and expands González-Fernández and Schmitt by contributing (a) the investigation of an additional L1 learner group; (b) the by-group as well as multigroup comparisons of EFL learners' vocabulary knowledge and acquisition sequence; and (c) a robust method for the empirical validation of the preliminary findings across different EFL learner populations that will serve as foundation for further examinations of acquisition orders of word-knowledge components.

The study addresses the following research questions:

RQ1: How well do EFL learners from two dissimilar L1 backgrounds (Spanish and Chinese) know various word-knowledge components? Does this knowledge differ by group?

RQ2: Is it possible to outline an order of acquisition of word-knowledge components that stays consistent across EFL learners of different L1 backgrounds when explored individually and collectively?

METHODOLOGY

Participants

Participants consisted of 314 EFL learners (254 females and 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%, adopted from González-Fernández and Schmitt's (2020) data), and 170 (54.1%, primary data) were L1-Chinese learners. They were recruited as volunteers by the author in Spain, China, and the United Kingdom, and comprised undergraduate and postgraduate students as well as professionals in various fields. Participants had a learning history in L2 English of between 1 and 26 years ($M = 11.81$, $SD = 3.90$), and most of them (76.8%) were studying English formally at the time of the study in different settings, including EFL lessons and English as medium of instruction. Following the implicational scaling requirement that participants need to have different proficiency levels, the dataset includes a representative sample of learners with a range of proficiencies 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 their estimated vocabulary level in English was collected through a compound score of the 2000 (2 K), 3000 (3 K), 5000 (5 K), and 10,000 (10 K) word-frequency sections of the vocabulary level test (VLT; Schmitt, Schmitt, & Clapham, 2001). The participants averaged 64.5% across the four sections (61.7% for the L1-Chinese participants and 68% for the L1-Spanish sample, internal consistency Cronbach's alpha = .83 and .92, respectively), which indicates that their overall vocabulary level was generally good. The two subsamples differed in their knowledge of most sections (F (12, 301) = 46.73, $p < .001$), except the 3 K frequency band (L1-Chinese: 2 K = 92.4%, 3 K = 74.0%, 5 K = 63.9%, 10 K = 16.4%; L1-Spanish: 2 K = 82.9%, 3 K = 74.3%, 5 K = 71.3%, 10 K = 42.9%). It should be noted that the 2 K band comprises items from both the 1 K and 2 K bands, which might explain why this was the only band being mastered (i.e., achieving at least 87% knowledge, Schmitt et al., 2001), although only by the Chinese

learners. Overall, results indicate that Chinese learners exhibited better knowledge of higher-frequency words, while the Spanish learners outperformed the other group in their knowledge of lower-frequency words.

Measures

To assess learners' knowledge of vocabulary components, González-Fernández and Schmitt's (2020) comprehensive battery of eight vocabulary tasks was adopted. This approach enables the direct comparison of results and thus the validation of their word-knowledge scale. The test battery follows Nation's (2022) description of word knowledge, which is understood in the conventional manner as mastery of various knowledge types for each individual word. It assesses explicit written recognition and recall knowledge of a set of 20 words across four different vocabulary knowledge components[4]: form–meaning link, derivatives, collocation, and multiple meanings. The 20 target words comprise: *mean, close, hard, development, season, bank, challenge, character, fresh, bright, broad, employ, distinction, charm, terminal, fulfill, grate, redeem, draught*, and *indent*. They provide the greatest opportunity to concurrently test the four word-knowledge components by meeting the following criteria: (a) range of frequencies (1 K-9 K) to account for the various proficiencies of participants, but with more words sampled at higher-frequency bands for representativeness; (b) multiple meanings, with at least three meanings as different from each other as possible (both polysemous and homonymous); (c) at least three derivative forms for the most frequent meaning; and (d) different parts of speech, to have a representative list of words. Thirty-five percent ($n = 7$) of the target words were Spanish–English cognates, and there were no loan words with Chinese (more details in Appendix A in the supplementary materials).

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 (Cheng & Matthews, 2018; Pellicer-Sánchez & Schmitt, 2010; Peters, 2016; Webb, 2005). The advantages of this approach are threefold: (a) it allows us to investigate how much an individual may know about each specific word, and thus, to explore overall word knowledge as described by most lexical frameworks (Nation, 2020); (b) it ensures that each component is being measured in a consistent and balanced

[4] Measuring all the components specified in Nation's (2022) framework is impractical. This test assesses four components that offer a representation of the three main constructs of word knowledge: form, meaning, and use. None of the 20 target words in the test battery appeared in the VLT.

manner in terms of the characteristics of the target words (Spencer et al., 2015), enabling comparability across components; (c) it controls for the effect that differential knowledge and features of the words assessed in each measure may have on the knowledge and implicational scale of lexical components (González-Fernández, 2022).

This approach, however, involves showing the specific set of words in multiple contexts and tasks, and thus, some practice effect between tests may occur. The test battery was purposefully designed to minimize the influence of this effect. González-Fernández and Schmitt ensured that the targeted derivatives, meanings, and collocations did not appear in the test before the task in which they were measured, and thus could not be answered based on information from previous tasks. In addition, the recall tests were administered before their recognition counterpart, and the various sections of the VLT were inserted between some tasks to minimize any possible memory of the target words from previous tests. González-Fernández (2022) further investigated the potential existence of practice effect in this test battery by means of Cochran's *Q* test with McNemar's post hoc analyses. She found that very few participants (3.58% on average) achieved mastery (established at a minimum of 75% task accuracy) of a typically complex test (e.g., derivative recall) without having demonstrated mastery of a generally easier and better-known task (e.g., form–meaning recognition). This finding indicates that, although some degree of practice effect in repeated testing circumstances is probably unavoidable (Nation & Webb, 2011), González-Fernández and Schmitt's efforts to control for and keep to a minimum the potential practice effect and overlap in their test battery were largely successful. This increases our confidence that the influence of one test on the others is minimal in the current study.

The complete test battery (available as Supplementary Material) exhibited high construct reliability for the participants in the present study (Composite Reliability = .94 for the L1-Chinese participants, .98 for the L1-Spanish participants and .96 for the total group of EFL learners). The individual tasks are concisely described below, and a sample item for each is provided. A more detailed account of each measure is offered in Appendix B, and the scoring methods and rationale for their selection is available in the supplementary materials of González-Fernández and Schmitt (2020).

Written form-recall knowledge of the form–meaning mapping. A fill-in-the-blank task where participants were asked to recall the English form of a word, given its most frequent meaning in a contextual sentence written in the participants' L1 (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"). Minor misspellings of the target item (e.g. *redem* for *redeem*) were accepted as correct. Cronbach's alpha [α] = .91 for L1-Spanish learners and .86 for L1-Chinese learners.

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_____.

Since the target participants are EFL learners, the most frequent meaning of the target words in this test battery was established by consulting corpus-based English learner dictionaries (e.g., *Cambridge Advanced Learners' Dictionary*) as opposed to L1-English corpora, and corroborated by two experienced EFL teachers. This is because "[L1-] frequency has its limitations when predicting which words might be known by L2 learners" (Schmitt, Dunn, O'Sullivan, Anthony, & Kremmel, 2021, p.1). For example, while according to COCA the meaning of the word *terminal* as a noun is more frequent than as an adjective in L1-English, the latter was chosen as the primary meaning in this study because it is considered more frequent for EFL learners in the learner dictionaries and by the two experienced teachers consulted. This decision enhances the ecological validity of the test battery for the target population.

Written meaning-recognition knowledge of the form–meaning link. A multiple-choice task where participants had to select the correct meaning from four frequent single-word options plus an "I don't know" option included to minimize guessing (L1-Spanish α = .91; L1-Chinese α = .72).

It is the best *season*.

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

Written form-recall knowledge of derivatives. Participants had to write down the target word's derivative forms that were appropriate in four sentences written to constrain word class, or an *X* if the word did not exist in a specific word class (L1-Spanish α = .96; L1-Chinese α = .85).

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 _____.

Written form-recognition knowledge of derivatives. A multiple-choice task where learners were given eight derivative options for each target word, with one correct option for each word class (or *X* if appropriate; L1-Spanish $\alpha = .95$; L1-Chinese $\alpha = .84$).

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 _____.

Written meaning-recall knowledge of multiple meanings. An open-question task where learners had to write, in their L1 or L2, a description, translation or synonym of three meanings of each target word, given the word class and a hint for each meaning tested (L1-Spanish $\alpha = .94$; L1-Chinese $\alpha = .86$).

Season

(Noun= year)_____

(Verb= cooking)_____

(Noun= animals in season)_____

Written meaning-recognition knowledge of multiple meanings. Participants had to choose all the sentences in which the target word was being used with a correct meaning, among three sentences depicting the three meanings tested and two sentences representing invented meanings of the word (L1-Spanish $\alpha = .91$; L1-Chinese $\alpha = .85$).

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

Written form-recall knowledge of collocates. Given a short context in the L1, participants had to fill-in the blank in the English sentence with the appropriate collocate of the underlined target word. Sample item translates as "Peak season is when most people go on holiday" (L1-Spanish $\alpha = .90$; L1-Chinese $\alpha = .81$).

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*.

Written form-recognition knowledge of collocates. A multiple-choice task where learners had to select the appropriate collocate from four options to fit a given sentence (L1-Spanish α = .90; L1-Chinese α = .76).

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 pen-and-paper format to small groups of participants or individually, depending on availability. The tasks were administered one-by-one following the order of administration piloted and proposed by González-Fernández and Schmitt (2020) to minimize the effect of previous sections on subsequent ones. Keeping the original administration procedure also ensures the comparability of results across studies. This administration order was as follows: form–meaning link form recall → VLT 5 K/3 K → form–meaning link meaning recognition → derivatives form recall → derivatives form recognition → multiple-meanings recall → collocate form recall → VLT 10 K/2 K → multiple-meanings recognition → collocate form recognition. Each task began with the instructions for completion and an example illustrating how to respond to the items. Participants were reassured that it was normal not to know all the answers, and that they should not guess. They handed in each separate task before starting the next one. The procedure lasted between 2 and 3 h (time limit 3.5 h), and participants were allowed short breaks between tasks.

Analyses

Descriptive statistics and one-way between-groups multivariate analysis of variance (MANOVA) were performed to empirically compare the extent of knowledge of each vocabulary aspect by the L1-Spanish and L1-Chinese EFL learners (RQ1). To explore the acquisition order of

the various vocabulary components (RQ2), Implicational Scaling and Mokken Scaling were employed. Implicational scaling [aka Guttman scaling (Guttman, 1944)] tests whether systematic hierarchical and implicational relationships exist between linguistic features, in this case the word-knowledge aspects. It measures consistency across individual learners' rank orders of knowledge rather than simply group data and thus validates whether the word-knowledge aspects can be hierarchically ordered across participants. This makes it a useful tool for the study of developmental systematicity in language (Rickford, 2002).

Implicational scaling examines learners at different stages in their vocabulary knowledge and thus can act as a proxy of general lexical development. According to implicational scaling, aspects that are less-known by most learners are considered more difficult to master by them, and what is systematically more difficult for learners is expected to appear later in the acquisition process (Buyl & Housen, 2015). Therefore, implicational scaling affords a basis for predicting which aspects learners would find easier or more difficult, providing a model of the anticipated order of acquisition[5] of word-knowledge components by a specific language community.

Implicational scales are tested by examining the distribution of mastered/unmastered vocabulary aspects for each individual participant, and arranging the various aspects in a hierarchical scale according to learners' general level of mastery, and thus relative difficulty. In accordance with previous guidance from implicational scaling research in SLA (Buyl & Housen, 2015; Hatch & Lazaraton, 1991), two operational criteria with different strictness levels were implemented in this study to consider a word-knowledge aspect mastered: 75% and 80% test accuracy (see González-Fernández & Schmitt, 2020 for the rationale for adopting these two accuracy levels). Thus, *mastery* and *mastered* in this study refer to achievement of these two accuracy thresholds for each vocabulary aspect. If the same scale remains when applying the 75% and 80% accuracy criteria, the scale gains statistical power.

Two goodness-of-fit measures were used to test the validity of the scale and the scalability of the word-knowledge aspects: Guttman's *Coefficient of Reproducibility* (C_{rep}) and *Coefficient of Scalability* (C_{scal}), respectively. C_{rep} shows how well we can predict the performance of a participant from their position in the matrix. The minimum reproducibility value of the C_{rep} to consider a scale reproducible, and thus valid, is .90 (Guttman, 1944; Hatch & Lazaraton, 1991). The C_{scal} reflects the strength of the aspects as an implicational scale, indicating whether

[5] Although predicted difficulty does not equate to actual development, the features of implicational scaling and its exploration of learners from different proficiency levels and their behavior individually and by group can provide a reasonable estimation of the expected acquisition order (Rickford, 2002).

the aspects are scalable and unidimensional. The variables are considered scalable if the C_{scal} is above .60.

To verify and support the findings from implicational scaling, additional Mokken scaling analyses were run (Van Schuur, 2011). Mokken analyses provide supplementary measures of scale homogeneity (Scale *H*; i.e., unidimensionality of the items) and reliability (*Rho*). A minimum *H* value of 0.3 is needed for a scale to be considered valid, with $H \geq 0.5$ indicating a strong scale, and *Rho* values of ≥0.7 represent a reliable scale (Sijtsma & Molenaar, 2002). The analyses were run in *R* (version 3.6.2) using the package *mokken* (version 3.0.6; Van der Ark, 2012).

RESULTS

Correlational analyses among the word-knowledge aspects were conducted by language group (L1-Chinese and L1-Spanish) and for the entire sample (see Appendix C). The results confirmed that the word-knowledge components were highly interrelated in both groups independently and combined ($r = .58–.87$ for the whole sample).

Vocabulary Test Scores

Regarding the first research question, Table 1 presents the mean percentage scores for the word-knowledge measures by language group and for the whole sample. Each group knew, on average, a minimum of half the items for each of the word-knowledge aspects, with

TABLE 1
Vocabulary Test Scores (%)

		FM Recall	FM Recog	Deriv Recall	Deriv Recog	MM Recall	MM Recog	Collo Recall	Collo Recog
L1 Chinese	*M*	53.15	79.38	51.57	65.82	50.29	64.90	57.21	77.32
$n = 170$	*SD*	11.99	9.53	10.26	10.74	10.00	11.62	15.38	12.50
	Range	10–85	50–100	28.8–83.8	27.5–88.8	25–80	28.3–91.7	20–100	30–100
L1 Spanish	*M*	53.61	82.81	51.64	61.08	49.69	70.76	60.66	79.69
$n = 144$	*SD*	18.91	13.79	20.86	19.90	16.12	15.18	19.05	17.39
	Range	10–95	40–100	5–91.3	18.8–97.5	6.7–90	31.7–96.7	15–95	25–100
Total sample	*M*	53.36	80.96	51.60	63.65	50.01	67.60	58.79	78.41
$N = 314$	*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

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.

scores ranging 49.7%–82.8% for the Spanish learners and 50.3%–79.4% for the Chinese learners. Table 1 also illustrates that some aspects were better known than others. In both groups, the best-known aspect overall was form–meaning recognition, while the worst-known aspect was multiple-meanings recall. Similarly, the recognition knowledge of each component exhibited a higher mean score than its recall counterpart across both language groups (between 60% and 80% accuracy in the recognition aspects, and 50% to 60% in the recall aspects for the entire sample).

To determine whether the two language groups differed significantly in their vocabulary knowledge, the results of the two subsamples across the various word-knowledge components were compared by means of one-way MANOVA. The test showed a significant effect of language group on the performance in the word-knowledge components: Pillai's Trace $V = 0.28$, $F(8, 305) = 14.51$, $p = .000$, partial $\eta^2 = 0.28$. Tukey's HSD post hoc tests with Bonferroni correction (resulting significance level $p < .007$) revealed that only the knowledge of multiple-meanings recognition was significantly different between the two language groups ($F(1,312) = 14.98$, $p = .000$) with a small-to-medium effect size ($\eta^2 = .05$). The mean scores for this aspect (Table 1) indicate that Spanish learners outperformed the Chinese learners in their recognition knowledge of multiple meanings. Non-significant language group differences were found on the rest of word-knowledge aspects (p between .008 and .97).

Implicational Scaling for L1-Chinese Learners

To answer the second research question, implicational scaling analyses were first conducted for the L1-Chinese learner group independently. The eight word-knowledge aspects were arranged in a hierarchical order according to their relative distribution of scores, from better known by learners to worse known. Learners were ranked relative to each other according to their performance across the aspects (from having mastered the greatest number of aspects to the least) and how well it matched the general order of aspects by the entire sample.

Following the stricter accuracy criterion, which required participants to achieve an 80% correct-answer threshold to consider an aspect mastered (see Analyses section), the resulting implicational scale was (from easier to more difficult):

Form–Meaning link meaning recognition
> *Collocation form recognition*
> *Multiple−Meanings meaning recognition*
> *Derivative form recognition*
> *Collocation form recall*
> *Form–Meaning link form recall*
> *Derivative form recall*
> *Multiple−Meanings meaning recall.*

The statistics in Table 2 show a very good fit of the 80%-accuracy scale. The .95 C_{rep}, which refers to the number of total responses that fit the implicational pattern (Hatch & Lazaraton, 1991), suggests that if an L1-Chinese learner could, for example, recall one of the word-knowledge aspects, we can predict that around 95% of the time that person would know all four recognition aspects. The C_{scal} represents the strength of the aspects as an implicational scale. The .67 C_{scal} shows that the implicational order shown in this scale is robust, so we can conclude that the word-knowledge components are scalable for the L1-Chinese learners and the implicational relationships between them are strong. Mokken analyses (see Appendix D) corroborated and supported the validity of the implicational scale at the 80% accuracy criterion. The same order of word-knowledge aspects was retrieved, and the fit indices (Table 2) indicated that the vocabulary aspects were highly homogenous ($H = .62$), confirming that they belong to one underlying construct (i.e., vocabulary knowledge, see also González-Fernández, 2022), and that the reliability of the scale is high ($Rho = .75$). Thus, a valid, robust, and statistically reliable order of acquisition of word-knowledge aspects was found for the L1-Chinese EFL learner group employing a strict 80% accuracy threshold.

TABLE 2
Implicational Scales' Goodness-of-Fit

		C_{rep}	C_{scal}	Scale H (SE)	Rho
	Acceptable fit	*≥.90*	*≥.60*	*≥.3/≥.5*	*≥.7*
L1-Chinese learners ($n = 170$)	80%-accuracy scale	.95	.67	.62 (.07)	.75
	75%-accuracy scale	.94	.60	.65 (.05)	.79
EFL learners ($N = 314$)	80%-accuracy scale	.94	.67	.75 (.03)	.86
	75%-accuracy scale	.94	.70	.79 (.03)	.88

C_{rep}, coefficient of reproducibility; C_{scab} coefficient of scalability; Scale H, homogeneity; Rho, reliability.

The validity of this scale was also examined under a slightly more lenient criterion: 75% accuracy. This cutoff point was considered the lowest level for which word-knowledge mastery could still be claimed (González-Fernández & Schmitt, 2020). The implicational scaling analysis following this threshold retrieved a pattern of components that was valid ($C_{rep} = .94$) and scalable ($C_{scal} = .60$), and Mokken analyses also confirmed the strong homogeneity (*Scale* $H = .65$) and reliability (*Rho* = .79) of the 75% scale (Table 2). In this scale, however, the distribution of components presented a slight variation as compared to the 80% scale, with the derivative recognition and multiple-meanings recognition aspects swapping places in the order. This alteration indicates that these two aspects are similar in the difficulty they pose to this group of learners, and thus, their rank order might interchange depending on the strictness of the accuracy criterion adopted. Yet, a comparison of the goodness-of-fit indices shows a considerably worse scalability for the 75% scale than the 80% scale ($C_{scal} = .60$ vs .67). This coefficient designates how confident we can be that the implicational scale exists, indicating that the ordering of word-knowledge aspects specified in the 80% accuracy scale is more robust and implicational. In addition, the 80% accuracy threshold can arguably be regarded as a more suitable criterion of mastery for the vocabulary components.

Consequently, the 80%-accuracy scale was considered the best representation of an order of acquisition of word-knowledge aspects for the L1-Chinese group. This implicational scale presents an identical ordering of aspects to that found by González-Fernández and Schmitt (2020) for their L1-Spanish learners. This result provides validity to their original scale and indicates that there seems to exist a reliable order of acquisition of word-knowledge aspects that is consistent across two distinct EFL learner populations when examined independently.

Implicational Scaling for EFL Learners

To investigate whether the implicational scale of word-knowledge aspects found for the groups independently remains consistent when the two EFL learner populations are analyzed collectively, further examinations were conducted with the whole sample (L1-Spanish and L1-Chinese learners combined). The implicational matrix for the 80%-accuracy scale in the entire sample is available in Appendix E.

These concurrent analyses resulted in an identical order of word-knowledge aspects across both accuracy criteria (80% and 75%) for the entire EFL learner group, which resembled exactly the one found for L1-Chinese learners independently in the strictest accuracy

criterion (80%; previous section). The scale obtained very good fit indices across both the 80% and 75% thresholds (Table 2), indicating the validity (C_{rep} = .94) as well as scalability (C_{scal} = .67 and .70, respectively) of the implicational scale. Supplementary Mokken analyses (Appendix D) validated the same distribution of aspects and showed that the scale has very strong homogeneity (H = .75–.79) and reliability (Rho = .86–.88).

In sum, these results show a consistent order of acquisition of word-knowledge aspects for the whole sample collectively and for the L1-Chinese learners independently. This order also concurs with the scale found by González-Fernández and Schmitt (2020) for L1-Spanish learners. Taken together, the study provides further evidence for the scalability of the vocabulary components in L2 English, indicating that a valid and reliable order of acquisition of the word-knowledge aspects might exist for EFL learners overall.

DISCUSSION

EFL learners' Knowledge of Lexical Aspects

This study examined the knowledge of eight word-knowledge aspects across EFL learners from two distinct L1 backgrounds (Spanish and Chinese). The two EFL groups demonstrated an overall similar degree of knowledge of the various vocabulary aspects. They were found to differ only in their recognition knowledge of multiple-meanings, in which the Spanish learners outperformed the Chinese learners. As has been previously claimed (e.g., Chen et al., 2012; Elgort, 2013), it is likely that the commonalities between Spanish and English derived from their cognate status might have given the Spanish learners a certain advantage in recognizing multiple senses of words; however, this cognate status could not counteract the additional complexity of recalling meanings compared to simply recognizing them (Peters, 2016). The study suggests that achieving recall knowledge of words is so complex for EFL learners that language similarities are not enough to support the extrapolation of recall knowledge from the L1 to the L2, thus behaving similarly to learners from non-cognate languages. Interestingly, the cognate advantage in the recognition of multiple meanings was not found in learners' knowledge of the other vocabulary components. This suggests, first, that the Chinese learners were not disadvantaged by being tested on some Spanish-English cognate words (although future studies could control for this effect), and second, that cognateness alone is not sufficient to lead to significant differences in EFL learners' overall word knowledge.

This indicates that developing general vocabulary mastery requires very advanced knowledge (Webb, 2009) and mental connections that may not be simply explained by metalinguistic knowledge alone. It is possible that formal instruction compensated for this effect. In the current sample, more learners in the Chinese group than in the Spanish group were studying English formally at the time the test took place (87.1% vs. 64.6%), which might have helped balance this cognate influence. This finding does not indicate that the cognate advantage does not exist for other aspects of word knowledge, but rather that enough training and exposure to the language (potentially with a focus on specific word-knowledge components) might counterbalance the cognate facilitation, at least as evidenced in offline written measures.

More importantly, the results indicated that the Spanish and Chinese EFL learners exhibited the same rank order in their knowledge of the various vocabulary aspects. This finding lends empirical evidence for the expectation of certain components of word-knowledge (i.e., form–meaning) being better known than others (i.e., derivatives) by L2 learners in general (Barcroft, 2002; Cheng & Matthews, 2018). Significantly, the study offers an overall pattern of knowledge of lexical components that extends to EFL learners from diverse L1s, and which allows us to compare the relationships and behavior of the word-knowledge aspects relative to each other. This pattern is discussed in detail in the following section.

General Sequence of Acquisition of Word-Knowledge Aspects

Understanding how the multiple aspects involved in vocabulary knowledge relate to each other and develop across various learner populations is crucial to establish a theory of L2 vocabulary acquisition, and to systematize the instruction of word-knowledge aspects (Schmitt & Schmitt, 2020). The present study provides an empirically supported, cross-linguistic implicational scale of word-knowledge components that remains constant for EFL learners of distinct L1s.

According to this hierarchical order of acquisition, the various word-knowledge aspects are known to different degrees and learnt incrementally by L2 learners. Recognition of the form–meaning link was the aspect mastered by the greatest number of learners, and thus is the easiest aspect in the implicational scale and expected to be acquired earlier by EFL learners. This finding provides evidence for the previously unsupported claim that form–meaning recognition is the first aspect of words to develop for L2 learners (Cheng & Matthews, 2018; Laufer & Goldstein, 2004; Pellicer-Sánchez &

Schmitt, 2010). Its position as the first aspect in the scale also reinforces the assertions that this component is the most important for vocabulary acquisition (Nation, 2020), as is the one that enables learners to start making use of language (i.e., retrieving the meaning of a word when encountered in reading; Laufer & Goldstein, 2004). Interestingly, this order indicates not only that knowledge of the form–meaning link is the starting point of vocabulary learning, but also that it is a prerequisite for the rest of components to develop, and thus, for attaining general lexical competence in an L2.

Importantly, the scale also allows us to compare how the word-knowledge components behave relative to each other. Collocational knowledge is known to be challenging to achieve for L2 learners, especially in recall mastery (Pellicer-Sánchez, 2017; Peters, 2016). This study found collocation knowledge to appear before that of multiple meanings and derivatives. This suggests that, while still demanding for learners, collocations does not seem to be the most complex aspect to develop in an L2 when compared to other less explored aspects, specifically if knowledge of only one recurrent collocate of each target word is required (a typical operationalization of this component in lexical research). In addition, both language groups demonstrated better recall mastery of collocations than recall of the form–meaning link. It seems possible that the similarity between these two tasks and the different frequency range of the forms that learners had to recall in each of them (i.e., 1 K-9 K for the form–meaning task and 1 K-3 K for collocates) made the collocation-recall task easier for the learners. This suggests that learners' recall knowledge of a frequent collocate of words is relatively good, and that this knowledge seems easier than recalling the form of lower-frequency target words, emphasizing the complexity of word-form recall particularly in lower-frequency bands (Barcroft, 2002; Chen & Truscott, 2010). Thus, the present study expands our understanding of L2 collocational knowledge by illustrating the behavior of collocations with high-frequency collocates in relation to other word-knowledge aspects.

Regarding derivatives and multiple meanings, the implicational scale shows that they were the most demanding components for the EFL learners, and thus can be expected to take longer to master both in recall and recognition knowledge. This finding lends support to previous uni-component studies which suggest that knowledge of derivatives (Barcroft, 2002; Nagy, Diakidoy, & Anderson, 1993) and knowledge of the meanings of words (Hoshino, 2015; Wolter, 2009) appear relatively late even for L1 speakers, and could be the last stages in L2 vocabulary acquisition. Thus, to achieve mastery of these complex aspects, learners would require significant time, rich training and extensive exposure to the L2.

Finally, the implicational scale highlights the fundamental distinction between recognition and recall knowledge in L2 vocabulary acquisition. It demonstrates that EFL learners of diverse L1s find it easier to achieve recognition mastery for all the aspects than recall mastery of any aspect tested in this study. This finding supports previous affirmations that recall mastery lags behind recognition mastery in an L2 (Laufer & Goldstein, 2004; Pellicer-Sánchez & Schmitt, 2010), and offers empirical evidence that this is the case not only within an individual component (e.g., derivatives form-recognition > derivatives form-recall) but also across components (e.g., derivative form-recognition > form–meaning form-recall). Importantly, the scale also confirms that lexical recall and recognition knowledge develop along a hierarchical sequence, and reveals that attaining recognition mastery of various components is necessary before learners can achieve recall mastery. This is because recall knowledge is more advanced, and thus contains and requires recognition mastery within it (Schmitt, 2014). Therefore, for learners to attain recall mastery of words, and thus be able to produce vocabulary, they would need to master sufficient recognition knowledge of various aspects first. This outcome highlights the great complexity of moving from recognition to recall mastery in an L2, indicating that this process requires considerable time and exposure to the L2 (Schmitt, 2019; Webb, 2009).

Overall, the findings in this study indicate that the acquisition of L2 word knowledge, like that of grammar or morphology, is determined by certain systematicity, where students of different L1 backgrounds with various proficiencies and from diverse learning contexts (e.g., EFL and naturalistic) seem to build L2 word-knowledge competence following a consistent pattern (at least for the target items and measures in this study). The cross-linguistic implicational ordering offers a clearer picture of how some vocabulary components interplay with each other, and it is an important step forward toward a better understanding of vocabulary acquisition in second languages.

PEDAGOGICAL IMPLICATIONS

The implicational scale validated in this study contributes a blueprint of expected lexical acquisition that can aid teachers to best prioritize and sequence the presentation of various word-knowledge features in the EFL classroom (Nation, 2020). Specifically, the scale indicates that vocabulary instruction should initially focus on getting L2 learners to recognition mastery of various word-knowledge aspects. To this aim, practitioners should ensure that learners are exposed to as much receptive/recognition vocabulary input as possible from the

beginning of instruction (from various sources such as reading, listening, or viewing [Nation, 2022]), and that this input includes the target words in different sentences and contexts so that various kinds of word knowledge appear. For example, while focusing on teaching form–meaning link recognition, this exposure to rich input would help learners consolidate knowledge of this aspect while beginning to develop intuitions of other aspects (e.g., collocation or derivative recognition). In addition, exposure to L2 input can be beneficial for learners to develop not only word-recognition knowledge but also the ability to recall the form and meaning of words, although to a lesser extent (Pellicer-Sánchez & Schmitt, 2010; Puimège & Peters, 2019). Once recognition knowledge of various aspects is achieved, learners will be better prepared to make the move toward the more complex recall knowledge of words. Recall knowledge is essential for robust word learning (Stewart et al., 2024). If the aim of vocabulary instruction is for our learners to use the words, pedagogical approaches should also be directed at this higher level of knowledge. Teachers can help learners progress from recognition to recall knowledge by actively teaching vocabulary through recall and productive learning activities, such as generative tasks that require using words in context (Lee, 2003; Webb, 2005). These productive tasks can consolidate and support the development of further recognition knowledge as well as help attain recall knowledge of words.

The hierarchical ordering also corroborates that the form–meaning mapping component should be the first target of deliberate teaching in the early stages of vocabulary learning (Cheng & Matthews, 2018). Once some knowledge of form–meaning has been achieved, teachers should gradually and incrementally move the focus of instruction to other more complex components of word knowledge (e.g., collocations and derivatives), contingent on learners' proficiency and the learning purpose. Multiple meanings and derivatives (particularly in recall mastery) appear to be more relevant for learners who aim at developing high linguistic accuracy and competence in the L2. It is worth noting, however, that the scale might not represent a one-size-fits-all solution. It is possible that under certain learning/teaching conditions the word-knowledge aspects require different degrees of attention which might impact the acquisition sequence. Thus, teachers should consider the proposed scale alongside the learning situation/purpose to decide when/how to present the word-knowledge components to their students.

Importantly, the implicational scale provides further evidence of the incremental learning of L2 vocabulary knowledge (Schmitt, 1998), where the various aspects of word knowledge build upon each other in a hierarchical manner. For learners to develop the lexical depth

required to use the target words appropriately in any given context (Schmitt & Schmitt, 2020), L2 vocabulary instructors and textbooks should provide learners with sustained and continued training and exposure to the word-knowledge aspects over time. This training should offer repeated encounters with the various types of knowledge of lexical items, presenting them in varied contexts and scenarios in each new encounter with the word. This repetition and recycling will be important for consolidation as well as elaboration and enhancement of the word-knowledge aspects (Webb, 2007). For example, if our learners are ready to benefit from the instruction of multiple meanings, presenting the vocabulary on multiple and rich reading tasks would help them achieve recognition knowledge of multiple-meanings, while at the same time revising and consolidating previously acquired aspects (e.g., form–meaning and collocations).

LIMITATIONS AND FUTURE DIRECTIONS

The current study expands our understanding of L2-English vocabulary knowledge and validates a preliminary order of acquisition of word-knowledge aspects by extending it to learners of two distinct L1s individually and collectively. Yet, the article presents some constraints that lead to new interesting research questions. Firstly, the internal consistency for the form–meaning recognition and collocation recognition measures was lower than would be expected (<.80), particularly for the L1-Chinese learners ($\alpha = .72$ and 76, respectively). This indicates greater variation in learners' scores for these measures, which is likely a result of learner's different attitudes toward and use of the "I don't know" option in these tests. Although this option was employed as an effective means to minimize guessing in multiple-choice tests (Zhang, 2013), evidence also shows that it can discourage some learners from reporting partial knowledge of the target words and, thus, have an impact on how learners are ranked (e.g., Stoeckel et al., 2019; Stoeckel, Bennett, & McLean, 2016). Future research could explore how removing this option influences learner behavior in these tests, and whether it impacts the implicational scale. The internal consistency values of the measures in this study also show that the Spanish participants as a group performed more consistently across the different tasks, supporting prior claims that language measures can function differently for different learner populations (Byrne, 2016). Thus, researchers and teachers need to consider that the specific characteristics of different L2 learner populations may lead to somewhat variable test performance. Despite this variation in test behavior, the implicational scale remained consistent across both

groups, indicating the robustness of the order of acquisition of word-knowledge components.

Secondly, in any research, the words and types of components targeted, and the measures employed to assess them, can influence the findings. Moving forward, research should assess different words and components, or employ alternative test formats, to see whether/how the implicational pattern varies. For example, given the need to assess multiple word-knowledge aspects and the length of this procedure, it was not practical to target more than 20 items in the present study. While this is an improvement compared to previous multicomponent studies (Webb, 2005, 2007 and Chen & Truscott, 2010 targeted 10 items), future studies could assess knowledge of more than 20 target words, sampling more items per frequency band, as this would provide a more comprehensive representation of overall lexical knowledge. Also, the finding that the collocation-recall task was easier for learners than the form–meaning-recall task highlights how measures may impact the results. Although the instruments were carefully constructed and piloted, there are many other ways in which these vocabulary aspects could have been tested. Further research could measure collocational knowledge using a different test format (e.g., L1-L2 translation), or by assessing lower-frequency collocates or various collocates for each target word, not only in their core meaning but also for other meanings of the word. This would help better understand the relationship between different levels of collocational knowledge and other word-knowledge aspects. Alternatively, other aspects such as spelling or associations could be assessed, focusing not only on explicit but also implicit knowledge (Godfroid, 2020), while keeping at least a portion of the components assessed in the current study for comparison.

Finally, the implicational scale found in this study represents the word-knowledge patterns of EFL learners from two distinct linguistics backgrounds, different learning contexts (i.e., EFL lessons and English as medium of instruction) and varied learner proficiencies (beginner to advanced). But to be confident about its generalizability, the findings need replicating with other L1 learner populations and in different learning settings (Long, 1990). It is possible that specific instruction or learning conditions can influence the sequence of acquisition. Future research should explore this possibility, for example by accounting for participants' learning history, or via experimental studies that apply certain teaching treatments (e.g., intentional or incidental, input flooding or pushed-output) on some word-knowledge aspect across various target words and compare the hierarchy of knowledge of aspects prior and post treatment. By conducting this type of replication and extension studies, we can continue to model the acquisition and development of word knowledge in second languages,

providing researchers, teachers, and testers with a framework for how to teach and test L2 vocabulary knowledge (Schmitt, 2019: 262).

ACKNOWLEDGEMENTS

I would like to thank the four anonymous reviewers for their insightful comments on earlier versions of this manuscript, which helped improve the article.

CONTRIBUTION AND FUNDING

None to report.

CONFLICT OF INTERESTS

The author declares none.

THE AUTHOR

Beatriz González-Fernández is a Senior Lecturer (Associate Professor) in Applied Linguistics and TESOL at the University of Sheffield, UK. Her research interests include the conceptualization, acquisition, and teaching of vocabulary in second/foreign languages. She has published in relevant journals in the field, including *Applied Linguistics* and *Studies in Second Language Acquisition*.

REFERENCES

Andersen, R. W. (1978). An implicational model for second language research. *Language Learning*, *28*(2), 221–282.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*(2), 323–363.

Buyl, A., & Housen, A. (2015). Developmental stages in receptive grammar acquisition: A Processability theory account. *Second Language Research*, *31*(4), 523–550.

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

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, *31*(5), 693–713.

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.

Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the vocabulary size test. *Language Testing*, *30*(2), 253–272.

Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics*, *18*(1), 4–22.
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). London: Routledge.
González-Fernández, B. (2022). Conceptualising L2 vocabulary knowledge: An empirical examination of the dimensionality of word knowledge. *Studies in Second Language Acquisition*, *44*, 1124–1154.
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.
Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, *9*(2), 139–150.
Hatch, E. M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House Publishers.
Hoshino, Y. (2015). Development of a test to measure knowledge regarding multiple meanings of basic words. *JLTA Journal*, *18*, 78–91.
Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, *21*(2), 202–226.
Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436.
Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System*, *31*(4), 537–561.
Long, M. H. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly*, *24*(4), 649–666.
Lubliner, S., & Hiebert, E. H. (2011). An analysis of English–Spanish cognates as a source of general academic language. *Bilingual Research Journal*, *34*, 76–93.
Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, *68*(2), 321–391.
Nagy, W. E., Diakidoy, I.-A. N., & Anderson, R. C. (1993). The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives. *Journal of Literacy Research*, *25*(2), 155–170.
Nation, P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 15–29). London: Routledge.
Nation, P. (2022). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.
Nation, P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.
Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*(3), 381–402.
Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*(1), 31–55.
Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113–138.
Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*(4), 424–438.

Rickford, J. R. (2002). Implicational scales. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 142–167). Oxford: Blackwell.
Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, *48*(2), 281–317.
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., Dunn, K., O'Sullivan, B., Anthony, L., & Kremmel, B. (2021). Introducing knowledge-based vocabulary lists (KVL). *TESOL Journal*, *12*(4), e622.
Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
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.
Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory* (Vol. *5*). London: Sage Publications Ltd.
Spencer, M., Muse, A., Wagner, R. K., Foorman, B., Petscher, Y., Schatschneider, C., … Bishop, M. D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing*, *28*(7), 959–988.
Stewart, J., Gyllstad, H., Nicklin, C., & McLean, S. (2024). Establishing meaning recall and meaning recognition vocabulary knowledge as distinct psychometric constructs in relation to reading proficiency. *Language Testing*, *41*(1), 89–108.
Stoeckel, T., Bennett, P., & McLean, S. (2016). Is "I Don't know" a viable answer choice on the vocabulary size test? *TESOL Quarterly*, *50*(4), 965–975.
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*, 1–14.
Trofimovich, P., Gatbonton, E., & Segalowitz, N. (2007). A dynamic look at L2 phonological learning: Seeking processing explanations for implicational phenomena. *Studies in Second Language Acquisition*, *29*(3), 407–448.
Van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software*, *48*(5), 1–27.
Van Schuur, W. H. (2011). *Ordinal item response theory: Mokken scale analysis*. London: Sage.
Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52.
Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65.
Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, *40*(3), 360–376.
Wolter, B. (2009). Meaning-last vocabulary acquisition and collocational productivity. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners: Papers and perspectives in honour of Paul Meara* (pp. 111–127). Bristol: Multilingual Matters.
Zhang, X. (2013). The I don't know option in the vocabulary size test. *TESOL Quarterly*, *47*(4), 790–811.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendices A–E. Supplementary Materials.

Data S1. Complete test battery.