Vocabulary knowledge is a complex construct that involves the acquisition of multiple word knowledge components (Henriksen 1999; Read 2000; Nation 2013; Author xxxx). However, most of our current understanding about this construct derives from studies that have assessed only one type of word knowledge, especially the form–meaning link (Melka 1997; Milton and Fitzpatrick 2014a). As a consequence, the construct of vocabulary knowledge as a whole is still largely unexplored, and it is unclear how the different word knowledge components are acquired and fit together (Author xxxx; Milton and Fitzpatrick 2014b). Paul Meara noted the absence of an overall theory of vocabulary acquisition in 1983, and despite the large amount of vocabulary research in the last decades, this is still the case in 2018. An obvious reason for this lack of a general theory of vocabulary acquisition is that researchers have generally not attended to the multidimensional nature of vocabulary in any great detail (although see a number of exceptions below), and have typically not explored the interrelationships between the word knowledge components they did study. This article is an initial step towards addressing this gap, by measuring L2 learners' knowledge of four word knowledge components (both in recognition and recall), and comparing the results in an attempt to begin modelling the relationships between the various components.

Background

The Components approach to vocabulary knowledge

This conceptualization of the overall knowledge of a word[1] divides word knowledge into its component parts. Richards (1976) is usually credited with the first detailed components list, which was further developed by Paul Nation in 1990. Nation's (2013) list is the most comprehensive, and the one referred to by most current vocabulary scholars. It includes nine different components (often referred to as types of *word knowledge*), each broken into receptive [R] and productive [P] mastery (Figure 1).

[FIGURE 1 NEAR HERE]

Although Nation's list provides the most exhaustive description of word knowledge to date, it does not specify the relationships between the components. This leaves important questions unanswered, such as the relative contribution of the various components to the vocabulary knowledge construct (e.g., does the form–meaning link explain most of the variance in vocabulary?) and whether some components tend to be acquired before others (e.g., are a word's derivative forms typically learned before its collocations?).

The list's very breadth is also a disadvantage, in that it is very complex to apply in research (i.e., it is virtually impossible to measure everything). For this reason, researchers have typically studied the components individually, e.g. form (Milton and Hopkins 2006); the form–meaning link (Laufer and Goldstein 2004); derivatives (Schmitt and Zimmerman 2002; Kieffer and Lesaux 2012b); collocations (Peters 2016); polysemy (Verspoor and Lowie 2003); and associations (Fitzpatrick and Izura 2011).

Although these single-component studies are informative for the component in question, they provide only limited insight into vocabulary knowledge as a whole, and how the various components fit together (Li and Kirby 2015). This has led a number of scholars (e.g., Author xxxx; Webb 2005; Milton and Fitzpatrick 2014a) to encourage the measurement of multiple components concurrently (including their receptive and productive masteries) with a battery of tests. However, given the complexity involved, there have been relatively few scholars who have taken up this challenge. Their studies, which address at least three word knowledge components, are reviewed in the next section.

Multicomponent vocabulary studies

One of the first studies to explore knowledge of multiple word knowledge components concurrently was Schmitt (1998). He conducted a longitudinal study into the vocabulary development from exposure of four English as a second language (ESL) learners in university settings over an academic year. He measured the productive knowledge of four word knowledge components of 11 words: written form, derivative knowledge, associative knowledge, and knowledge of multiple meanings. He found that these four components of word knowledge were interrelated somehow, and improved more or less in a parallel manner. In a similar cross-sectional study, Chui (2006) studied the knowledge of 20 words from the Academic Word List (AWL) (Coxhead 2000) across 186 English as a foreign language (EFL) learners for four components: word class recognition, meaning recall, collocation recognition, and derivative form production. She found that word class recognition and meaning recall were better known by the students, and therefore might be acquired earlier than the knowledge of the productive derivative forms or the recognition of collocations.

While the above two studies explored the acquisition of vocabulary components without any intervention, some other multicomponent studies have included an exposure/instructional intervention. Webb (2005, 2007a) examined how 66 and 121 (respectively) Japanese EFL students acquired 10-20 nonwords from various exposures. He used an extensive battery of written tests to measure five components of word knowledge, both productively and receptively: orthography (written form), form–meaning link, syntax (syntagmatic associations), grammatical functions (word class) and (paradigmatic) associations. He found that orthography was the component that enjoyed the highest gains, and that the rest of the components varied over the different exposures to the words. Overall, the different components developed in a parallel manner, but at slightly different rates.

Following Webb (2007a), Chen and Truscott (2010) investigated the acquisition of four word knowledge components by 72 Taiwanese university students (orthography, parts of speech, and associations (both receptively and productively), and form–meaning link (receptively only)), and found that increasing repetitions lead to better knowledge in all the different components, although the gains in knowledge varied depending on each component.

These multicomponent studies suggest that the various components are interrelated, but that some appear to be acquired before others. However, there is simply not enough evidence to indicate an acquisition order of these components, or to specify the nature of their interrelationships. The results of previous studies have been summarized in Appendix 1, which shows that the degree of (gained) knowledge of the various components was different across studies, even when the exact same measures were used (Webb 2005, 2007a, 2007b, 2009). Thus, the results are inconsistent regarding what appears to be acquisition orders based on the mean scores. This is partly because word knowledge components have been conceptualized and tested inconsistently. Different studies measure different components, or they measure the same components but with different tests (see Appendix 2 for a summary of the tests employed in multicomponent studies). For example, knowledge of word parts has been measured with tests of part-of-speech knowledge (Webb 2007a; Chen and Truscott 2010), inflectional and derivational suffixes (Schmitt and Meara 1997) or only derivational suffixes (Chui 2006), or identifying the base of multi-morphemic words (Li and Kirby 2015). Thus, the lack of consensus on how to best test a specific component makes it difficult to compare the results. Also, researchers have applied different learning techniques (e.g., Webb 2005, 2007b; Pellicer-Sánchez and Schmitt 2010), which could have enhanced specific types of word knowledge leading to different performance on the various components in the different studies (Webb 2005, 2009), or no learning technique at all (e.g., Schmitt 1998; Chui 2006). However, the main reason for the inconsistency in results is that these studies were not designed with the purpose of exploring an acquisition order in mind, and only examined the strength of the relationships between word knowledge components using simple correlational and mean scores analyses. In order to explore the possibility of an acquisition order of word knowledge components, specific statistical analyses are required, and the design of the study needs to accommodate to the features of such analyses.

In sum, although the above studies describe some types of word knowledge, they have limitations which restrict their ability to inform how multiple knowledge components of words develop in relation to each other. Some of these limitations are: testing only receptive or productive mastery (Schmitt 1998; Tannenbaum *et al.* 2006), or testing both but inconsistently across components (Chui 2006; Chen and Truscott 2010; Li and Kirby 2015); measuring multiple components, but for different target words (Milton and Hopkins 2006; Kieffer and Lesaux 2012a; Li and Kirby 2015); relatively few words tested (Schmitt 1998; Webb 2005; Chen and Truscott 2010); interpretation of mean scores and correlations only (Webb 2005; Chui 2006; Webb 2007a; Chen and Truscott 2010); low numbers of participants (Schmitt 1998; Webb 2005); and use of nonwords (Webb 2005, 2007a).

These limitations suggest that the time is ripe for a study which concurrently measures multiple word knowledge components, at two levels of sensitivity, and using more sophisticated statistical procedures which can better show the relationships between these components. The current study follows this approach in order to explore the following research questions:

1. To what degree are components of vocabulary knowledge interrelated?
2. As individual words are learnt, are some components of word knowledge typically acquired before other components?
3. What is the best way of conceptualizing the relationships between the various word knowledge components?

Methodology

Participants

The participants included 144 Spanish-speaking learners of English as L2 (102 females, 42 males), whose age ranged from 18-65 years (*M* = 25.25, *SD* = 8.04). They were recruited as volunteers in Spain and the UK, and all had formally studied English for a minimum of four years. Since the use of Implicational Scaling requires the subjects to have different proficiency levels, we aimed for a population of learners with a range of proficiency in English, from beginners to advanced. Almost half of the participants (49.3%) reported themselves as having an intermediate general proficiency in English, 38.5% considered themselves advanced users of English, and only 13.2% rated themselves as beginners in English. A compound score of the 2,000 (2K), 3,000 (3K), 5,000 (5K), and 10,000 (10K) most frequent English words on the Vocabulary Levels Test (VLT) (Schmitt, Schmitt, and Clapham 2001) was used as an estimate of the learners' vocabulary size. The participants averaged 68% overall across the four levels, (*M* raw score = 81.4/120, *SD* = 20.15), with 71% correct answers in 5K level (*M* raw score = 21.35, *SD* = 5.38) and 43% in the 10K (*M* raw score = 12.85, *SD* = 5.46). This suggests that their overall vocabulary knowledge was relatively good on average, although we were also successful in recruiting lower-proficiency learners, some of whom obtained only 22-35% correct answers.

Target Words

Twenty words were selected which provided the greatest opportunity to test the four targeted word knowledge components (see below), based on the following criteria:

- Range of frequencies (1K-9K), to account for the different proficiencies of the participants and meet requirements of the Implicational Scaling analysis.
- Different parts of speech, in order to have a representative list of words.
- Multiple meanings, with at least three senses as different from each other as possible. This usually resulted in homonymous (semantically-unrelated) senses, but sometimes also in polysemous (semantically-related) senses, with many target words including both.[2]
- At least three derivative forms for one of the meaning senses.
- A percentage of Spanish-English cognates representative of the proportion in general English. Some indications are that the percentage is around 34-37% (Lubliner and Hiebert 2011). Therefore, the target words included 35% cognates (n = 7) and 65% non-cognate words (n = 13).

The resulting 20 target items included: *mean, close, hard, development, season, bank, challenge, character*, *fresh, bright, broad, employ, distinction, charm, terminal, fulfil, grate, redeem, draught*, and *indent.* These target words have ecological validity in that they represent the different types of items students may encounter (e.g., cognate words[3], words from different word classes, different frequencies, and different types of multiple meanings).

The same twenty words were tested across the four word knowledge components, because we were interested in exploring how knowledge of individual words develops. This approach has been used in previous depth of vocabulary knowledge studies (e.g., Schmitt 1998; Webb 2007a; Chen and Truscott 2010), and is a standard way of exploring the development of depth of knowledge of individual words.

Measures

There are many ways in which word knowledge components can be conceptualized and measured, as Appendix 2 illustrates. One main consideration is how to measure what is typically referred to as *receptive* vs. *productive* knowledge. Following Schmitt (2010), the terms *receptive* and *productive* knowledge are employed as skills-based definitions of language ability, meaning, respectively, knowing a lexical item well enough to extract communicative value from speech or writing, or well enough to produce it when it is needed to encode communicative content in speech or writing. Since it is difficult to measure truly receptive and productive knowledge in this sense, we have taken the approach used by most previous studies and have employed some form of *recognition* or *recall* test format to assess knowledge of the different components. In this study, *recognition* was understood as a type of knowledge when the learner is able to recognize and select correct word knowledge information about a target item from a number of given options, and *recall* as being able to retrieve word knowledge information from memory (without options) after some stimulus is given.

In our study, we prioritized written test formats that have been used in previous published research, and were compatible with the purposes of the study. In this section, we briefly illustrate each measure, and give an example. See Appendix 3 for a more complete description of each format, the rationale for its selection, and the complete test battery.

Form recall knowledge of the form–meaning link

This test followed a fill-in-the-blank format, where participants were asked to recall the L2 form, given the L1 meaning (e.g., Laufer and Goldstein 2004; Webb 2005). The participants were presented with a context in Spanish setting the situation and meaning of the target word (e.g. the example below for *season* translates as “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”).

El verano es el mejor periodo del año para mí, porque me gusta mucho el calor y poder ir a la playa.

It is my favorite s_______________.

Meaning recognition knowledge of the form–meaning link

This test utilized the format that is perhaps the most widely-used in research assessing this type of knowledge: a meaning recognition multiple-choice test (e.g., Laufer and Goldstein 2004; Webb 2005).

It is the best <u>*season*</u>.

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

Form recall knowledge of derivatives

This task took the format that has been used in previous research analysing productive knowledge of derivatives (e.g., Schmitt and Zimmerman 2002; Saigh 2015). In this task, participants were asked to write down the derivative forms (if existent) of the target word that were appropriate in four sentences written to constrain meaning and word class.

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

The design of this test was based on the format used by Saigh (2015), and adopts a multiple-choice task with multiple answers. The learners were presented with eight different derivative options for each target word, with one correct option for each word class.

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

This task involved a written open-question format, in which students were assessed on three senses of each target word. They were given the target word, plus the word class, and a hint about each of the three meanings tested. After each clue, they had space to write, in their L1 or L2, a translation, a synonym, a description, a definition, or a sentence in which the specific meaning tested was used clearly.[4]

Season

(Noun= year)____

(Verb= cooking)____

(Noun= animals in season)____

Meaning recognition knowledge of multiple meanings

This measure was adapted from a format used by Li and Kirby (2015). In this format, each target word was presented in five different sentences, with a different sense 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. The participants were told that there was a minimum of one and a maximum of three correct sentences in each item, in order to minimise blind guessing.

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

Learners' form recall knowledge of collocations was assessed using a test format based on previous research (Author xxxx; Peters 2016). Participants were given a short context in Spanish (the example below means "Peak season is when most people go on holiday"), and had to fill in the English sentence gap with the appropriate collocate given the first letter. The sentences were written so that there were no direct translations which the participants could use to answer the items without knowing the collocates.

En temporada alta es cuando más gente se va de vacaciones.

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 task employed a multiple-choice format, based on previous research (e.g., Chui 2006; Webb, Newton, and Chang 2013). Students were presented with a sentence in which the target word was underlined and they had to choose the appropriate collocate from four options, plus an 'I don't know' option.

There are always more tourists in ___________ *season*.

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

Procedure

The test battery was administered in pen-and-paper format, and the different sections (i.e., each individual test) were numbered and clearly distinct. The order of administration of sections was designed and piloted to minimise the effect of previous sections on subsequent ones,[5] with the VLT (Schmitt *et al.* 2001), breaking up the components tests: 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. The 2K VLT level was always placed at the end of the battery to put the easiest VLT section towards the end of a long test. Participants were given the option to choose from English or Spanish instructions and explanations, depending on their confidence with the language.

Each section started with specific instructions on how to complete it and examples illustrating how to respond to the items. Participants handed in each individual section before starting the next one, to minimise cross-contamination between sections. The battery of tests was administered to small groups of participants or to participants individually, depending on their availability.

Analyses

Correlations were carried out in order to obtain an overall view of the degree of interrelatedness between word knowledge components. Then *Implicational Scaling* (henceforth IS) was employed to estimate and analyse the difficulty in acquisition of the various word knowledge components. IS (aka Guttman Scaling (Guttman 1944)), allows the establishment of systematic hierarchical relationships between variables, and can be considered a proxy for systematicity in language (Rickford 2002). Thus, it can help make predictions about how the various word knowledge components are acquired. In order to support the findings from IS, we also ran a Mokken analysis with the Program MSP5.0 (Molenaar and Sijtsma 2000). We used a 75%-80% correct criterion in these analyses, and henceforth, the terms *mastery* and *mastered* refer to achievement of this threshold.

After the IS and Mokken analyses, *Structural Equation Modelling* (SEM) was used to provide an explanation of the relationships between word knowledge components that has inferential power. SEM is a set of statistical techniques that are theory-driven and follow a confirmatory approach to check the validity of a previously hypothesized model of relationships between variables (Byrne 2016). SEM was chosen because of its advantages over other analyses. It allows for the specification and analysis of latent, unobservable constructs by means of multiple observed indicators. These latent constructs are more reliable than any individual indicators, and less sensitive to the effects of the specific tasks used (Kieffer and Lesaux 2012b). Moreover, it allows for the specification of theoretical models that establish multiple relationships among several variables and examines the whole set of relationships among these variables simultaneously. Finally, it measures the strength of relationship between each path while taking into account all the other paths in the model. SEM produces a set of model fit indexes that shows how well the data fits the model (Kline 2016). See Appendix 4 for details on how these procedures were carried out.

In this study, SEM was used to examine the nature of the overall vocabulary knowledge construct. The most widely accepted conceptualization of vocabulary knowledge is Nation's (2013) framework of what is involved in knowing a word (Figure 1). According to this framework, vocabulary knowledge is seen as comprised by multiple word knowledge components, with each of these components being made up of both receptive and productive knowledge. We modelled this conceptualization as seen in Figure 2. This hypothesised model represents vocabulary as a general, underlying latent construct comprised by smaller sub-constructs (i.e., form–meaning link, derivatives, multiple-meanings and collocations) which, in turn, are represented by their recall [Recall] and recognition [Recog] masteries. By using SEM, we seek to empirically test this conceptualization.

[FIGURE 2 NEAR HERE]

Results

Descriptive statistics (Table 1) show that, on average, learners knew about half of the items for each test, and that the recognition mastery of an aspect always shows higher scores than its recall mastery. The wide gap between the minimum and maximum scores shows that our word selection was successful in obtaining a range of words which the participants knew to greater and lesser degrees.

[TABLE 1 NEAR HERE]

Correlations

In order to explore the interrelatedness of the components, a correlation analysis was run. Because all the main variables in this study were non-normally distributed, Spearman’s correlations were computed (Table 2). The correlations between word knowledge components lie between .700 and .945, showing a high degree of relationship. It is also interesting that knowledge of the various word knowledge components correlates very highly with vocabulary size, as indicated by the VLT Total score (.760-.895). This supports Author’s (xxxx) conclusion that vocabulary size and depth are inextricably related to one another, and that word knowledge components seem to be highly interconnected.

[TABLE 2 NEAR HERE]

Implicational Scaling

The eight word knowledge aspects[6] were arranged horizontally in a matrix according to difficulty, that is, ordered from better known to worse known (left to right). Participants were ranked vertically relative to each other according to their general performance. Using the criterion of 80% correct answers to consider an aspect mastered[7], the following implicational scale was observed (from easier to more difficult):

Form–Meaning link meaning recognition > Collocate form recognition > Multiple-Meanings meaning recognition > Derivative form recognition > Collocate form recall > Form–Meaning link form recall > Derivative form recall > Multiple-Meanings recall

The Guttman statistics show that the goodness-of-fit of this scale is very good. The *Coefficient of reproducibility* (C_{rep}) was .94, and the *Coefficient of scalability* (C_{scal}) was .73. The C_{rep} exceeds the minimum reproducibility value of .90, and thus we can say that a valid implicational scale exists (Guttman 1944). This means that knowledge of a higher aspect in the scale implies the knowledge of all lower ones (e.g., multiple-meanings recognition implies both collocate form recognition and form–meaning link recognition). Thus, according to this C_{rep}, if a participant can recall one aspect, it would mean that around 94% of the time, that person would know all four aspects at the recognition level. The C_{scal} reflects the strength of the components as an implicational scale, indicating whether the aspects are unidimensional, and therefore scalable. The data is considered scalable if the C_{scal} is above .60, although the higher the value of the C_{scal}, the more 'implicational' the scale is (Davidson 1987). The high C_{scal} (.73) indicates that the pattern of scalability is very robust and the aspects unidimensional.

The results from the Mokken analysis showed that the aspects had a very strong homogeneity (H = .79), confirming that the components form part of one underlying construct (i.e., vocabulary knowledge), and that the reliability of the scale is very high (*Rho* = .89). Thus, we can speak of a strong, statistically reliable order of difficulty of the word knowledge components for this sample.

Structural Equation Modelling

This section will provide a brief overview of the SEM analyses. For the validation of the measurement models and a full description of the analyses, see Appendix 4. The hypothesised model presented in Figure 2 (Model 1) was fitted to SEM using the maximum likelihood robust estimator, with the *lavaan* package, version 0.5-23.1097 (Rosseel 2012) in the software *R Studio* (R core Team 2017, version 3.4.1; R Studio Team 2016, version 1.0.153). Following the commonly employed fit indexes and guidelines conventions (e.g., Hu and Bentler 1999; Brown 2015), a good model fit is generally indicated by a nonsignificant chi-square (χ^2), chi-square/*df* ratio (χ^2/df) of between 1 and 3, comparative fit index (CFI) and Tucker-Lewis index (TLI) of >.95, standardised root mean square residual (SRMS) < .08, and root mean square error of approximation (RMSEA) of < .05 for good fit and < .08 for acceptable fit. Figure 3 shows the strength of the relationships between word knowledge components and vocabulary knowledge as described in Model 1. Although this model showed a good overall fit with the data ($\chi^2 = 26.43$, $df = 16$, $p = .05$; CFI = .99; RMSEA = .07), the regression coefficients (i.e., paths between Vocabulary Knowledge and the four word knowledge components) were all very high (β = .94-.98), suggesting lack of discriminant validity (Kline 2016) and affecting the significance of some paths. These β coefficients indicate that all these components are highly intercorrelated, and that they can be better understood as a single construct. That is, the regressions between them were too high to claim that they were different sub-constructs of vocabulary knowledge. Thus, this model is not well supported by the data, and cannot be considered a good representation of vocabulary.

[FIGURE 3 NEAR HERE]

Since Model 1, which divided vocabulary knowledge into multiple components, was not supported by the data, some modifications of this hypothesised model were needed. We reconceptualised the model based on the results from the IS, which showed that recognition and recall knowledge in general behaved quite differently from each other. Therefore, we respecified the model considering recognition and recall aspects as individual direct indicators of the general Vocabulary Knowledge construct. Moreover, based on the high correlations shown in Table 2, we hypothesised that the recognition and recall aspects of the same component (e.g., Derivative Recall and Derivative Recognition aspects) would correlate. This revised model (Model 2) is illustrated and analysed in Figure 4.

[FIGURE 4 NEAR HERE]

[TABLE 3 NEAR HERE]

As shown in Table 3, the results of model evaluation show that all the model fit indexes reached and exceeded the commonly accepted fit thresholds, supporting the suitability of the revised model, and thus, the validity of the construct. Therefore, Model 2 appears to be a good representative of vocabulary knowledge, based on our data[8].

In general, this model suggests that the recall aspects tend to contribute to vocabulary knowledge slightly more than their recognition counterparts, with the exception of the Form–Meaning Recall/Recognition pair. Nevertheless, the factor loadings are very high for all word knowledge aspects, suggesting that all of them contribute in a similar manner to, and are explained by, a unique factor: Vocabulary Knowledge. In sum, all word knowledge aspects make a large contribution to the explanation of the Vocabulary Knowledge construct, which demonstrates they are all essential components of knowing vocabulary.

Discussion

This study explored the overall nature of written vocabulary knowledge, using the most extensive multicomponent test battery to date with a large number of participants (N = 144). What did this large amount of data tell us about lexical knowledge? In some cases, it confirmed common assumptions about vocabulary which have been made until now with little or no empirical evidence. In other cases, it lead to surprising results, which challenge typical views of vocabulary.

The interrelatedness of word knowledge components

Ever since Nation first presented his initial listing of eight word knowledge components in his 1990 book, people have thought about the components as separate entities. Nevertheless, there have been claims that the components are interrelated in some way (e.g., Webb 2008). Our data demonstrates that these claims were correct. We found significant correlations between the various aspects ranging from .700 to .945, indicating that there are clearly strong and comprehensive connections between all of the word knowledge aspects we measured. This is interesting because it seems that no aspect is learned in a way that is detached from the other aspects. Vocabulary learning is incremental, and there seems to be a considerable amount of parallel learning occurring in the process (Webb, 2007a). This suggests that knowing one aspect facilitates the learning of other aspects. Yet the acquisition rate is not even among the aspects, since the implicational scale shows that some aspects tend to be mastered before others.

There has also been a great deal of debate concerning the relationship between breadth (size) and depth (quality) of vocabulary knowledge. Depth has been conceptualized in many ways (Author xxxx), and our study adds to this discussion by demonstrating the close relationship between vocabulary size (as measured by the VLT) and the various word knowledge components. Our correlations are closely clustered between .760 and .895. This means that the shared variance between vocabulary size and word knowledge 'depth' varies between 58% and 80%. The VLT measures knowledge of the form–meaning link at the form recognition level (Author xxxx), so one might assume that the VLT scores would correlate most closely with the form–meaning recognition and recall tests. We find that the VLT actually correlates more strongly with the Derivative, Multiple-Meanings, and Collocate measures. This suggests that vocabulary size is strongly related to knowledge of the various word knowledge components in general, and not just the ones which have a similar testing format (i.e., focusing on the form–meaning link). There is also tendency for the size-aspect correlations to be somewhat higher for the recall measures than the recognition measures. But overall, it seems like the greater your vocabulary size, the better you will know the words across all aspects.

An acquisition order of word knowledge components

The fact that the various word knowledge aspects are closely interrelated, however, does not mean that they are mastered simultaneously. Intuitively, some components would clearly seem easier (e.g., form–meaning) or more difficult (multiple-meanings) than others, but until now, there has been little hard evidence indicating the order of acquisition of the various components. Our study is the first to our knowledge to suggest a statistically-reliable implicational scale of written vocabulary knowledge. Previous studies focusing only on single word knowledge components have shown that recall knowledge seems to be generally more difficult than recognition for an individual aspect (e.g., the form–meaning link (Laufer and Goldstein 2004), collocations (Peters 2016)). But previous research was inconclusive as to whether all the recall aspects are more difficult than all recognition aspects, or whether some recall aspects can be easier than some recognition aspects. For example, Pellicer-Sánchez and Schmitt (2010) found that the two recall components measured (word class and meaning) were learnt after the two recognition components (meaning and spelling). Conversely, Pigada and Schmitt (2006) found that a recall component (spelling) was easier than some recognition aspects (grammatical knowledge).

The key finding from this study is that recognition knowledge of all four of the components we measured was mastered (at least in terms of the measures, criteria, and limitations of our study) before any type of recall knowledge for those components. In other words, recall mastery seemed hard for L2 learners across a range of component types. Previous research has usually shown this (e.g., Pellicer-Sánchez and Schmitt 2010; Van Zeeland and Schmitt 2013), but typically for only individual components. For example, Laufer and Goldstein (2004) demonstrated that receptive knowledge (recognition) of the form–meaning link was easier than productive knowledge (recall), forming the following order (> = more difficult than): form recall > meaning recall > form recognition > meaning recognition. Our implicational scale involving recall and recognition knowledge of four different components builds on this research by attesting that the easiness of recognition before recall knowledge applies to all of the components we measured. This finding suggests that the most important distinction in vocabulary knowledge may not be between the word knowledge components themselves, but rather between the relative recognition and recall mastery of those components.

Nevertheless, there does seem to be a general ordering in the acquisition of the various written components. Some previous studies used specific learning techniques (deliberate and incidental), and the potential acquisition orders from those studies (Appendix 1) might be affected by the specific treatment technique employed (Webb 2005, 2009). Our study presents a general learning situation, where different kinds of learners (i.e., different in their proficiency levels) have learnt different vocabulary from different teaching and learning situations. Thus, we can conclude that when measuring a varied sample of words (different frequencies and cognate status) and not being situated in any particular learning environment, the written word knowledge components do seem to be ordered in how well they are known.

Overall, the Form–Meaning link was the best known, and generally appears before other components, as shown in earlier studies (Henriksen 1999; Tannenbaum *et al.* 2006; Pellicer-Sánchez and Schmitt 2010). However, this result seems to contradict other previous multicomponent research (e.g., Webb 2005, 2007a; Chen and Truscott 2010) which suggests that other components precede the acquisition of the Form–Meaning link. In Webb's studies, there was a tendency for knowledge of spelling and (sometimes) word class to develop before the form–meaning link, although this order varied according to number of encounters, treatment and/or the time on task. We did not measure spelling or word class knowledge, so we cannot comment directly on how they might fit into our acquisition order, other than to say that it is possible that some untested word knowledge components may prove to be easier than the Form–

Meaning link, i.e. the Form–Meaning link is not necessary the beginning of a complete acquisition scale.

Chen and Truscott (2010), found that the Form–Meaning link was the worst known aspect for very low frequency words. We also found that form recall of the Form–Meaning link was difficult for our learners for low-frequency words. In fact, this partially explains one of the surprising results of our study: that form recall of the Form–Meaning link proved more difficult than Collocate form recall. In the Form–Meaning link recall test, learners needed to spell the target word (from 1-9K frequency bands), while in the Collocation recall test they needed to provide the collocate, all of which came from within the 1-3K frequency bands. It appears that, overall, learners know the very frequent collocates to a form-recall mastery marginally better than they know the spelling of all the lower frequency target words. Therefore, we believe that our results confirm the conclusion of several scholars (e.g., Laufer 1998; Barcroft 2002) that form recall is trickier than is commonly thought, especially for lower frequency words.

Knowledge of form concerning the syntactically-constrained knowledge of word family members (as in the Derivatives measures), also seems difficult and is learned relatively late (Nagy, Diakidoy, and Anderson 1993; Chui 2006). Indeed, it is probably not surprising that Derivatives (both recognition and recall) are among the last aspects to be learned. This corroborates arguments by Barcroft (2002) stating that form knowledge is one of the most difficult components to acquire, and probably requires explicit teaching attention.

Collocations have traditionally been considered a difficult component, with many studies showing them as problematic for L2 learners (e.g., Webb *et al.* 2013; Peters 2016). Many might have expected that they would be further towards the difficult end of the scale. We interpret this finding not as showing that collocations are easier than believed before, but as showing that derivative and multiple-meanings knowledge are harder than previously thought. The cognate nature of Spanish may have also helped our participants. Elgort (2012) found that cognate status of words reliably increases the response accuracy of items in a bilingual (English-Russian) size test.

The results are probably also partially an artefact of the measures we used. We asked for only a single collocation, but for four derivative forms. Chui (2006) found that recognition knowledge of collocations was known almost at the same level as productive knowledge of derivatives, but she asked her learners to produce only one derivative, instead of our four. This highlights the fact that our implicational scale is based on our particular methods of measurement, and that different methods may produce somewhat different results. Nevertheless, we feel our measures are all valid and reliable, and it is impossible to assess various aspects with exactly the same formats in any case, e.g., measuring the form–meaning link entails one link, while measuring derivative knowledge entails measuring (at least) the four major word classes (noun, verb, adjective, adverb). Our study offers an initial attempt to tap into the word knowledge acquisition order, and only future research with different measures and learner populations will support/refute its generalizability. We suspect that although the acquisition order of components may change somewhat with different participants/measures, the recognition/recall distinction is likely to remain. Such a study with 170 Chinese learners of English is now being prepared (Author, in preparation).

Knowledge of multiple senses also seems to be a more difficult component to acquire. Previous research suggests the acquisition of even a single meaning is not straightforward. Van Zeeland and Schmitt (2013) report that meaning recall was acquired incidentally from listening after spelling and word class recognition. Wolter (2009) believes that meaning is a component generally learned late. Schmitt (1998) looked at the acquisition of multiple meanings and found that L2 learners can achieve knowledge of other word knowledge components without having mastered a word's multiple senses. It is not difficult to speculate why multiple-meanings knowledge is relatively problematic. Most pedagogical materials focus on single senses, and the amount of exposure required to come across less frequent senses is exponentially greater than the most frequent sense. For example, in a 100-condordance-line random sample of the COCA Fiction corpus, the 'money' sense of *bank* occurred 68 times, the 'riverside' sense 19 times, the 'collection' sense (*bank of clouds*) 12 times, and the 'tilting' sense (*banking airplanes*) only once.

Although our scale is statistically reliable, it is also potentially misleading in ways the reader needs to be aware of. The implicational scale indicates that the different word knowledge components follow a general difficulty order of acquisition. However, this does not mean that they are always strictly learned in sequence. The strong intercorrelations suggest that below the 75% criterion level, there may be some parallel learning accruing. For example, there is probably some degree of recall knowledge present before recognition knowledge is mastered at the 75-80% level. Also, individual words have their own idiosyncratic characteristics (e.g., cognateness, frequency) which can make some word knowledge components relatively easier or more difficult to acquire. Likewise, learners vary and may be better at picking up some components than others. Nevertheless, although the acquisition scale may not hold for any particular word or person, it represents group tendencies and is consistent across varying types of words, learning conditions, and learner proficiencies.

Conceptualizing vocabulary knowledge

Nation’s (2013) framework of word knowledge components has been very influential, and has helped researchers and practitioners conceptualize vocabulary knowledge for a generation. Our study enhances this framework by giving an indication of how the components relate to one another. While the IS analysis provides information about the difficulty order of the written word knowledge aspects, the SEM analysis shows how these components relate to each other as a whole vocabulary construct. The present study is the first to directly examine the relationships among the word knowledge aspects using latent variables. An advantage of using latent variables is that the relationships among the variables are examined free of measurement error and reliability of the measures does not affect the relationships (Tannenbaum *et al.* 2006). Thus, this analysis can provide an accurate representation of the relationships between the aspects of word knowledge.

Nation's (*ibid*.) framework suggests that receptive and productive knowledge are part of each word knowledge component, which in turn make up the greater vocabulary knowledge construct. However, when we modelled this conceptualization, understanding recognition as a more receptive end of the receptive/productive continuum and recall as relatively closer to the more productive end, we found that the four word knowledge components loaded on the Vocabulary Knowledge construct extremely strongly and almost identically. This is largely because all of the recognition knowledge aspects behaved similarly to each other, as we saw in the IS. The same was true for the recall aspects. This meant that no word knowledge component was known both in recognition and recall before another component. Rather, all of the recognition knowledge aspects were mastered before the recall ones. This lead to the four components being known to a very similar degree, and could not be discriminated by the model.

When we modelled the recognition and recall knowledge aspects as independent factors in Model 2, the fit was extremely good, indicating that recognition vs. recall knowledge was the key distinction in our study. This conceptualization suggests that perhaps the focus of pedagogy should be shifted towards pushing learners' knowledge from receptive towards productive mastery. Language teaching materials, if they focus on vocabulary at all, generally merely introduce new words, often as glosses of reading passages or as defined words in vocabulary boxes. The lexical descriptions or tasks are usually limited to the meaning of a word, aiming to help the learner recognize and understand the word in discourse (i.e., receptive knowledge). There is typically very little recycling of vocabulary, or the kind of exercises necessary to develop productive mastery of the word (Brown 2010). This study provides evidence that developing towards the productive end of the receptive/productive continuum is complex and requires more time. Productive knowledge of all components comes later in the learning process, and would seem a sensible target for pedagogy.

Limitations and Future Research

Although this study was carefully designed, there are some limitations that can be the starting point for future research. In the first place, the vocabulary knowledge construct is comprised of more components than we could include in this study due to practical constraints. Future research could usefully explore other components of word knowledge (both in oral and written mode), such as constraints on use, grammatical functions and associations. Previous research by Chen and Truscott (2010) showed that the receptive form–meaning link was more difficult than knowledge of orthography, part of speech and associations. Our study showed that receptive form–meaning link was generally known better than collocations, multiple-meanings and derivatives. Combining the information of the two studies gives some indication of the relationships between these seven components. Given the impracticality of measuring *all* word knowledge components concurrently, future studies should explore different combinations of components in order to build a composite picture of the overall word knowledge component constellation. Also, apart from measuring other vocabulary components, future studies exploring the acquisition order of vocabulary should aim to test more than twenty target words if possible, in order to obtain a wider language sample.

Another limitation is the fact that this study examines the relationships of word knowledge components in English by speakers of only one L1. This leaves us unsure of whether these results are universal, and therefore can be applied to any other EFL learners, or whether they are language-dependent. Therefore, replications of this study with other language populations would provide a more complete picture of the acquisition of vocabulary components in English as an L2. Moreover, the analysis of vocabulary knowledge in this study is based on cross-sectional examination of components. Thus, the conclusions regarding the evolution of the different word knowledge aspects are limited. In order to provide a clearer understanding of the development of each of these word knowledge components over time, longitudinal research is needed.

Finally, as in any research, the results are influenced by the way things are measured. Appendix 2 shows that there is no one valid way of testing word knowledge components. Although our measurement instruments were carefully constructed and piloted, there are many other ways in which the various vocabulary aspects could have been tested. For example, future research could account better for the great complexity of the multiple meanings construct and control for polysemous and homonymous meaning senses, in order to explore how this factor affects meaning acquisition. It is possible that with different measures, we might find different results, and this possibility needs to be explored. Only then can we begin to understand the degree to which our initial results are generalizable. Regardless, we believe that our study is a useful beginning step towards the development of an empirically-based general theory of vocabulary acquisition.

Notes

1. In this study we are interested in words that have the same written form and multiple meanings, disregarding the origin of the different senses. Therefore, the target words include both polysemes and homonyms. This basic 'multiple-meanings' conceptualisation does not take into account the linguistic distinctions between polysemy and homonymy. The distinction is interesting because polysemous (semantically-related) vs. homonymous (semantically-unrelated) meaning senses might have different degrees of difficulty for acquisition. Unfortunately, given all of the other word selection criteria that were required to tap into the various word knowledge aspects, it was not possible to control for polysemy/homonymy as a variable of interest in this study. We also chose the 'multiple-meanings' conceptualisation on pedagogical grounds, as learners come across both polysemes and homonyms in their studies, and so we thought it useful to include both in our study.

2. It is important to note that we are dealing with only the written forms of the words, where the spelling form remains identical for the different meaning senses even if the pronunciation may change for the various word classes (e.g., *indent* as a noun is pronounced /ˈin dent/, and as a verb /in ˈdent/).

3. When administering the same test battery to a non-cognate language population (Chinese), initial findings suggest that cognateness does not appear to affect the overall order of acquisition or the SEM relationships between word knowledge components (Author, in preparation). Also, although participants generally knew cognate words better than non-cognate words, the order of the mean percentage scores for the various word knowledge components for cognates was the same as the order for noncognates. The effects of other factors such as frequency, proficiency, and word length are being examined in a separate follow-up study.

4. In the test materials, we used the term *polysemy* instead of 'multiple meanings' because we thought that this would be the term most learners would be familiar with based on the Spanish cognate (*polisemia*).

5. Interviews were conducted to see if participants were able to use information from previous tests on later tests. They commented that, although they could remember that the target words had occurred before, they could not remember any information that helped them in subsequent tests. This suggests that our efforts to minimise the information interference were largely effective. Despite this, there might still have been some test effect, even if minimal, as this is potentially unavoidable when several tests are employed (Nation and Webb 2011).

6. In this paper, we use *component* to refer to the word knowledge components (e.g., derivatives, collocation) and *aspect* to refer to the separate recognition/recall dimensions of knowledge (e.g., derivative recall, derivative recognition, collocation recall, collocation recognition).

7. We used the 80% accuracy criterion typically found in IS research, but this cut-off point is admittedly arbitrary. We thought that 75% accuracy would be the lowest level for which word knowledge mastery could still be claimed, and so also tried this slightly more lenient criterion to see if the scale still obtained, which it did (C_{rep} = .95 and C_{scal} = .81). This suggests that the implicational scale remains regardless of the cut-off criteria used, which provides further evidence of the scalability of these items. The Mokken analysis confirmed the strong homogeneity and unidimensionality of the 75% scale (*H* = .85), and very high scale reliability (*Rho* = .92).

8. We cannot claim that Model 2 is the *only* valid statistical representation of vocabulary knowledge, but it is the model that best fit our data, with its particular measures and participants.

References

Author. xxxx.

Author. xxxx.

Author. xxxx.

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

Brown, D. 2010. 'What aspects of vocabulary knowledge do textbooks give attention to?,' *Language Teaching Research* 15/1: 83–97.

Brown, T. A. 2015. *Confirmatory Factor Analysis for Applied Research* (2nd ed.). The Guilford Press.

Byrne, B. M. 2016. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (3rd ed.). Routledge.

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

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.

Coxhead, A. (2000). 'A new academic word list,' *TESOL Quarterly*, 34/2: 213–238.

Davidson, F. 1987. *Guttman Scalogram Analysis*. Unpublished manuscript, University of Illinois at Urbana-Champaign.

Elgort, I. 2012. 'Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test,' *Language Testing* 30/2: 253-272.

Fitzpatrick, T. and **C. Izura.** 2011. 'Word association in L1 and L2,' *Studies in Second Language Acquisition* 33/3: 373–398.

Guttman, L. 1944. 'A basis for scaling qualitative data,' *American Sociological Review* 9/2: 139–150.

Henriksen, B. 1999. 'Three dimensions of vocabulary development,' *Studies in Second Language Acquisition* 21/2: 303–317.

Hu, L. and **P. M. Bentler.** 1999. 'Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives,' *Structural Equation Modeling: A Multidisciplinary Journal* 6/1: 1–55.

Kieffer, M. J. and **N. K. Lesaux.** 2012a. '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.

Kieffer, M. J. and **N. K. Lesaux.** 2012b. 'Direct and indirect roles of morphological awareness in the English reading comprehension of native English, Spanish, Filipino, and Vietnamese speakers,' *Language Learning* 62/4: 1170–1204.

Kline, R. B. 2016. *Principles and Practice of Structural Equation Modeling* (4th ed.). The Guilford Press.

Laufer, B. 1998. 'The development of passive and active vocabulary in a second language: Same or different?,' *Applied Linguistics* 19/2: 255-271.

Laufer, B. and **Z. Goldstein.** 2004. 'Testing vocabulary knowledge: Size, strength, and computer adaptiveness,' *Language Learning* 54/3: 399–436.

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

Lubliner, S. and **E. H. Hiebert.** 2011. 'An analysis of English–Spanish cognates as a source of general academic language,' *Bilingual Research Journal* 34/1: 76-93.

Meara, P. 1983. *Vocabulary in a Second Language, Vol. 1*. Specialised Bibliography 3. CILT.

Melka, F. 1997. 'Receptive vs. productive aspects of vocabulary' in Schmitt, N. and M. McCarthy (eds.): *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge University Press.

Milton, J. and **T. Fitzpatrick.** 2014a. *Dimensions of Vocabulary Knowledge*. Palgrave Macmillan.

Milton, J. and **T. Fitzpatrick.** 2014b. 'Introduction: Deconstructing vocabulary knowledge' in Milton, J. and T. Fitzpatrick (eds.), *Dimensions of Vocabulary Knowledge*. Palgrave Macmillan.

Milton, J. and **N. Hopkins.** 2006. 'Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners?,' *Canadian Modern Language Review* 63/1: 127–147.

Molenaar, I. W. and **K. Sijtsma**. 2000. *MPS5 for Windows. A Program for Mokken Scale Analysis for Polytomous Items*. iecProGAMMA.

Nagy, W. E., **I. N. Diakidoy** and **R. C. Anderson.** 1993. 'The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives,' *Journal of Reading Behavior* 25: 155–170.

Nation, I. S. P. 1990. *Teaching and Learning Vocabulary.* Newbury House.

Nation, I. S. P. 2013. *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press.

Pellicer-Sánchez, A. and **N. Schmitt.** 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.

Pigada, M. and **N. Schmitt.** 2006. 'Vocabulary acquisition from extensive reading: A case study,' *Reading in a Foreign Language* 18/1: 1–28.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

R Studio Team. 2016. *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA. Retrieved from URL http://www.rstudio.com/

Read, J. 2000. *Assessing Vocabulary*. Cambridge University Press.

Richards, J. C. 1976. 'The role of vocabulary teaching,' *TESOL Quarterly* 10: 77–89.

Rickford, J. R. 2002. 'Implicational scales' in Chambers, J. K., P. Trudgill and N. Schilling-Estes (eds.): *The Handbook of Language Variation and Change*. Blackwell.

Rosseel, Y. 2012. 'lavaan: An R package for structural equation modeling,' *Journal of Statistical Software* 48/2: 1–36.

Saigh, K. 2015. *Derivative Knowledge and the Development of an Affix List for Pedagogy*. Unpublished PhD thesis. University of Nottingham.

Schmitt, N. 1998. 'Tracking the incremental acquisition of second language vocabulary: A longitudinal study,' *Language Learning* 48/2: 281–317.

Schmitt, N. 2010. *Researching Vocabulary: A Vocabulary Research Manual*. Palgrave Press.

Schmitt, N. and **P. Meara.** 1997. 'Researching vocabulary through a word knowledge framework,' *Studies in Second Language Acquisition* 19/1: 17-36.

Schmitt, N., **D. Schmitt** and **C. Clapham.** 2001. 'Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test,' *Language Testing* 18/1: 55–88.

Schmitt, N. and **C. B. Zimmerman.** 2002. 'Derivative word forms: What do learners know?,' *TESOL Quarterly* 36/2: 145–171.

Tannenbaum, K. R., **J. K. Torgesen** and **R. K. Wagner.** 2006. 'Relationships between word knowledge and reading comprehension in third-grade children,' *Scientific Studies of Reading* 10/4: 381–398.

Van Zeeland, H. and **N. Schmitt.** 2013. 'Incidental vocabulary acquisition through L2 listening: A dimensions approach,' *System* 41/3: 609–624.

Verspoor, M. and **W. Lowie.** 2003. 'Making sense of polysemous words,' *Language Learning* 53/3: 547–586.

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. 2007a. 'The effects of repetition on vocabulary knowledge,' *Applied Linguistics* 28/1: 46–65.

Webb, S. 2007b. 'Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge,' *Language Teaching Research* 11/1: 63-81.

Webb, S. 2008. 'Receptive and productive vocabulary sizes of L2 learners,' *Studies in Second Language Acquisition* 30/1: 79–95.

Webb, S. 2009. 'The effects of receptive and productive learning of word pairs on vocabulary knowledge,' *RELC Journal* 40/3: 360-376.

Webb, S., **J. Newton** and **A. Chang.** 2013. 'Incidental learning of collocation,' *Language Learning* 63/1: 91–120.

Wolter, B. 2009. 'Meaning-last vocabulary acquisition and collocational productivity' in Fitzpatrick, T. A. (ed.): *Lexical Processing in Second Language Learners: Papers and Perspectives in Honour of Paul Meara*. Multilingual Matters.

Supplementary Data

The following appendices may be found in the Supplementary Data section of the online version of this article:

Appendix 1: Tentative order of acquisition of multicomponent studies

Appendix 2: Word knowledge aspects measured in multicomponent studies and test formats

Appendix 3: Test battery and rationale

Appendix 4: Implicational Scaling and Structural Equation Modelling procedures