THIS IS AN EARLIER VERSION OF THE MANUSCRIPT. FOR THE FINAL VERSION PLEASE CHECK THE JOURNAL WEBSITE:

https://doi.org/10.1017/S026144481800037X

Language Teaching-Replication Paper

Examining second language vocabulary growth: Replications of Schmitt (1998) and Webb & Chang (2012)

Ana Pellicer-Sánchez

Biographical note

Ana Pellicer-Sánchez is Associate Professor of Applied Linguistics and TESOL at the UCL Institute of Education. Her research centres around the teaching and learning of vocabulary in a second or foreign language, with a particular focus on learning from reading. Her recent research has made use of eye-tracking to explore the cognitive processes involved in vocabulary learning.

Abstract

There has been extensive research in the last 20 years on the effectiveness of different instructional interventions and learning conditions on the acquisition of vocabulary. However, very few attempts have been made to explore how vocabulary knowledge develops over time. This paper argues for the need to conduct more longitudinal studies on vocabulary learning and teaching and provides suggestions for important replication studies in the area. In particular, this paper calls for the replication of the studies by Schmitt (1998) and Webb & Chang (2012). Unlike most studies on vocabulary learning and teaching, these two follow a longitudinal approach and study vocabulary growth from two main perspectives, i.e. the

development of vocabulary depth and vocabulary breadth. The approximate replications suggested here would constitute an important contribution to the field of vocabulary learning and teaching.

1. Introduction

Vocabulary is an essential component of language proficiency and, thus, has featured high in the language learning and teaching research agenda. Numerous studies have shown the close relationship between learners' vocabulary knowledge and performance in different language skills. Vocabulary knowledge has been conceptualised in terms of two main dimensions: VOCABULARY SIZE (or breadth), referring to the quantity of words that learners know, and VOCABULARY DEPTH, alluding to the quality of vocabulary knowledge, i.e. how well those words are known (Anderson & Freebody 1981). Studies on vocabulary size have shown that second language learners need to know a large number of words in order to function in English, with estimates ranging from 3,000-4,000 word families for the comprehension of spoken discourse to 8,000-9,000 word families to understand a range of written texts (Nation 2006), and that learners often struggle to meet these targets (e.g. Dang & Webb 2014). Furthermore, research has shown that learning new vocabulary is more than just learning the connection between a new form and its meaning. A component approach to vocabulary tries to capture all the different aspects of knowing a word (Milton & Fitzpatrick 2014), with Nation (2001) providing the most comprehensive list to date of the different word knowledge components. Addressing these two main approaches, vocabulary studies in the last twenty years have focused on investigating effective ways to support learners in getting those larger vocabulary learning targets as well as to learn the different components of lexical mastery.

However, most vocabulary learning studies that examined the effectiveness of instructional interventions for the improvement of quantity and quality of vocabulary

knowledge have followed the typical pre-post-test design, focusing on the acquisition of a small set of items usually taught in one (or very few) session(s) and measuring learning immediately (or shortly) after the treatment. We need to be cautious in the interpretation of results from these one-shot design studies. Just because we observe that in one session a particular approach has led to the acquisition of 10 new words, we cannot assume that, if learners have three sessions a week, they will learn 30 words per week and 120 words every month. As Schmitt (2010) explains, the learning rate for a particular type of practice will probably not be maintained as practice increases. Similarly, different aspects of vocabulary knowledge will also need different amounts of time and exposure to develop and we cannot assume that the development experienced in a one-month treatment by several aspects of vocabulary knowledge will represent how lexical aspects grow over a longer time. Thus, results from these non-longitudinal studies cannot be used to make claims about how vocabulary knowledge develops over time.

Acknowledging the incremental and gradual nature of vocabulary knowledge implies that 'only research designs with a longitudinal element can truly describe it' (Schmitt 2010, p. 156). The incremental nature of vocabulary growth can be explored in relation to the two facets of vocabulary knowledge explained above. We can look at how learners increase their vocabulary size overtime, as well as how the different components of lexical mastery develop over time. Despite the clear benefits and implications that this better understanding of vocabulary growth has for language teaching, longitudinal investigations in vocabulary learning research are scarce.

Most existing studies exploring changes in VOCABULARY SIZE at different levels of proficiency and different stages of the acquisition process have followed a cross-sectional design (e.g. Milton 2006; Henriksen 2008; Milton 2008). Very few studies have tracked

vocabulary development of the same group of learners over time. These studies have usually been conducted in the study abroad (SA) context, examining the development of vocabulary during the period abroad and measuring vocabulary knowledge before and after the SA period (e.g. Milton & Meara 1995). To my knowledge, only one study, conducted by Webb & Chang (2012), has examined growth in vocabulary size with the same learners across several years.

Regarding VOCABULARY DEPTH, previous studies have suggested that certain aspects of vocabulary knowledge are learned before others (e.g. Schmitt & Meara 1997; Webb 2005; Pigada & Schmitt 2006) and interesting group differences have emerged in cross-sectional investigations of specific word knowledge components (e.g. Nagy, Diakidoy, & Anderson 1993). Very few studies have attempted to track the development of word knowledge components over a longer period of time and, with the exception of the study by Schmitt & Meara (1997), most studies have examined the development of a single learner. For example, Churchill (2008) tracked his own acquisition of the form and meaning of one word over three months. Similarly, Bell (2009) examined the development of knowledge of 17 items by one learner using essays written over 16 months. A wider range of components was examined by Fitzpatrick (2012). She explored the incremental changes in the vocabulary knowledge of one learner over the course of one year of overseas study. The development of five components was examined, using the Lex30 (a word association task), to tap into the learner's productive vocabulary knowledge. Her study showed that some aspects grew gradually (i.e. collocations and associations), whereas striking inconsistencies appeared in other components (i.e. word form, form-meaning connections, and orthography). The study by Schmitt (1998) was the first one to employ a procedure to look at the acquisition of a wider range of components (using different tests for each component) with more than one single learner.

In general, results from these existing studies have shown the 'complex and multi-dimensional nature of concurrent lexical growth' (Fitzpatrick 2012, p. 82). Our knowledge of this complex nature of vocabulary growth is still rather compartmentalized. We are still far from reaching a conclusion of how the many different aspects of lexical mastery that Nation (2001) identified develop over time in relation to one another and what vocabulary size figures we should expect at different stages of learning. This paper presents an argument for the approximate replication of two key studies on vocabulary growth that cover the two main dimensions of lexical knowledge, i.e. depth (Schmitt 1998) and breadth (Webb & Chang 2012). These two investigations are among the very few that have empirically explored vocabulary growth using a longitudinal design. Looking at vocabulary growth, both in terms of how many words learners acquire over time and how word knowledge components develop, has important implications for pedagogical practices and for the construction of a vocabulary learning model that is still missing, which would substantially contribute to the advancement of the field of vocabulary learning and teaching.

2. Replication Study 1: Schmitt (1998)

2.1. Background to the study

The aim of Schmitt's (1998) study was to examine the incremental acquisition of second language vocabulary over the course of a year. By tracking the acquisition of a small set of words, he was able to explore how the different components of vocabulary knowledge developed over time. Vocabulary research was booming in the late 90s. In Schmitt's (1998) words there was a 'virtual explosion of vocabulary studies' (p. 282). However, most of these studies focused on either the words that learners needed to know in order to operate in English (e.g. Goulden, Nation, & Read 1990) or the number of words that could be learned from reading (e.g. Huckin, Haynes, & Coady 1993) and from different types of explicit

instruction (e.g. Avila & Sadoski 1996). These studies were a crucial contribution to the advancement of vocabulary research but did not really explain how words were acquired over time. Importantly, vocabulary researchers at the time were advocating for the need to look at the acquisition of a range of lexical components in order to understand the interrelationships among the different components and to provide empirical evidence for the descriptive component lists available (e.g. Richards 1976; Nation 1990). However, judgements about words being learned in most experimental studies were based on the measurement of form and meaning. Only a couple of studies had made some initial attempts to look at the acquisition of other word knowledge aspects (e.g. Nagy, Diakidoy, & Anderson 1993; Schmitt & Meara 1997). The paucity of longitudinal vocabulary research, together with the need to adopt a multi-componential approach to vocabulary learning is what instigated the present study. Schmitt's study was the first one to examine how a wider range of word knowledge components of the same words developed over time, namely written form, associations, grammatical behaviour, and meaning. After twenty years of vocabulary research, Schmitt's (1998) study continues being unique. To my knowledge, no other studies have examined how knowledge of a range of lexical aspects of the same words develop over time in a group of learners.

This exploratory study followed the lexical development of three advanced learners of English who were doing a post-graduate course at an English university. The participants were initially measured on their knowledge of a set of eleven words. An important criterion in the selection of target items was that learners would need to be exposed to them during the course of their studies. Crucially, the target items needed to fall at various points of the acquisition continuum for participants, with some words well known, others partially known, and others unknown. This was confirmed in an initial pilot with 12 other students of similar characteristics. Eleven polysemous words from the University Word List (Xue & Nation

1984, reprinted in Nation, 1990) and from the Brown word list (Francis & Kucera 1982) were included in the study. Knowledge of four components (WRITTEN FORM, ASSOCIATIONS, GRAMMATICAL BEHAVIOUR, and MEANING*)* of the eleven target items was measured three times in the course of a year. Vocabulary knowledge was measured in one-on-one interviews in which participants were asked to: recall the written form of the target items presented orally (spelling test), provide three associates for each target word, recall the word class of the form and the three derivative forms, and recall the three meaning senses of the target words (first unprompted and then using prompts).

Results of the study showed that acquiring the different meaning senses of the target words was not easy, with the vast majority of meaning senses staying at that same level. However, in those cases where meaning knowledge improved, little forgetting occurred. Also, items known at the productive level were less prone to decay. Making the move from no knowledge to productive mastery was the most challenging type of growth. Knowledge of the correct form of the target words seemed to be easier. Any level of meaning knowledge almost always implied knowledge of the correct spelling. The associations provided by two of the three participants progressed and became more native-like with time, whereas for the third participant they remained at about the same level. Interestingly, meaning associations seemed to improve as knowledge of meaning senses increased, whereas they were not that closely related to grammar knowledge. Only one of the three students showed development in grammatical knowledge. Overall, the results of the study showed that different components progressed in slightly different ways, with some aspects improving more than others, and that the strength of the relationship of different word knowledge components seems to vary between different knowledge types. However, no specific developmental sequencing of word knowledge components was found.

2.2. Approaches to replication

This highly-cited study deserves replication because it was the first to look at the development of the different word knowledge components of the same words over time. Although some other similar attempts followed it (e.g. Churchill 2008; Bell 2009; Fitzpatrick 2012), the study by Schmitt (1998) is still the only one that looked at the vocabulary growth of a group of learners, as opposed to a single case study, and that used different tests to examine the development of a range of word knowledge components. Despite the significant contribution of this original study to the field of vocabulary research, it remains a very small-scale, exploratory investigation. A series of replications might provide more validity to the initial findings and confirm and expand the patterns found in this original study.

A first approach to close replication would be to keep the aim, research questions and methods employed in the original study but to collect data from a larger, homogeneous group of participants. As explained above, Schmitt's (1998) study reported the vocabulary development of three learners, measuring changes in the knowledge of eleven words at three points in a year's time. As Schmitt argued, participant attrition is likely to occur in this type of longitudinal studies. However, since the study only involved three sessions with each participant over the course of a year, having a larger population would still seem feasible. Adding more participants would allow researchers to perform not only descriptive (as it was the case in Schmitt 1998) but also inferential statistical analysis, providing more robust findings and confirming the degree to which the results might be generalized to other learners. For example, running a Two-way ANOVA with time (three time points) and lexical component (written form, associations, grammatical behaviour, and meaning) as independent variables and vocabulary scores as the dependent variable (provided that the assumptions of this test are met) would allow researchers to explore the main effect of time and type of word knowledge on vocabulary scores, as well as the interaction between these two factors. This close replication of the original study would allow researchers to look at average proportion

scores and compare them to the results of the original study, while at the same time adding statistical power to the original results through inferential statistics.

Several modifications in the assessment methods used in the original study should be considered for an approximate replication. The meaning senses in Schmitt's (1998) study were assessed first unprompted, simply asking participants to recall the meanings of each target word. Keywords were then used as a prompt to elicit the meaning senses that had not been provided in the unprompted phase. As Schmitt explains, this procedure attempted to tap into the receptive-productive meaning knowledge distinction. Receptive knowledge of the different meaning senses could also be examined by means of a multiple-choice test at the end of the procedure. This approximate replication would include the same unprompted, recall phase and the prompted recall phase that Schmitt used, but also an easier, multiple-choice recognition test that might be able to tap into smaller variations in the knowledge of meaning senses. Importantly, if we have these three levels of measuring meaning, all options in the final, multiple-choice items should be related to the key words provided in the prompted stage. Otherwise, the keyword might influence response behaviour in the recognition test. This modification in the measurement instruments would allow researchers to compare the two levels of receptive knowledge (prompted recall phase and multiple choice test) to the original prompted scores and provide a more detailed picture of receptive-productive knowledge. Another option might be to have just the unprompted recall test and the less demanding, multiple-choice meaning recognition test. A second methodological modification in this approximate replication would involve the addition of a second rater for the scoring of test results, and the consequent report of inter-rater reliability. Schmitt (1998) provides a very detailed explanation of how the different tests were scored. Although the scoring of some of the tests (e.g. grammatical knowledge) was more straightforward, as the

author acknowledges, the scoring involved a certain amount of subjectivity. Adding another rater would provide reliability to the original findings.

A further suggestion for approximate replication would involve some modifications in the design to be able to provide a more systematic examination of individual variation. The three students in Schmitt's (1998) study reported different strategies and approaches to vocabulary learning. For example, one of them reported having looked up the target words in the dictionary in between sessions, which seemed to support the acquisition of components like grammatical knowledge. Interesting individual variations were found among the three participants of the original study, some of which could not be discussed with the data collected nor the author's observations. Thus, one methodological addition in an approximate replication would be to include a more structured interview about learners' strategies and approaches to vocabulary learning. This would allow researchers to explain some of the differences observed among participants and verify the author's observations reported in the original study. In addition to strategies for learning, the original study by Schmitt could not provide any information about the type of exposure that learners had to the target items, which could be used to explain some of the patterns found. Churchill (2008) used a diary to record the learner's conscious encounters with the target word. A second modification might therefore involve adding this type of learner's diary in order to get a better picture of the type of exposure that learners have to the target items. The triangulation of the data from the vocabulary tests, the structured interview and the learners' diaries would allow us to get better insights of learners' strategies to vocabulary learning and of the amount and type of exposure that they had to the vocabulary items during the study, which would help to explain patterns of variability in the quantitative results. A final, fruitful area for close replication would be to modify the type of target items included in the study so that the learning gains for single words in the original study could be compared to the acquisition of formulaic

language. It is widely acknowledged that learning vocabulary involves learning items beyond the single word (e.g. Wray 2002; Schmitt 2004; Wray 2008; Wood 2010; Wood, 2015). In order to achieve high levels of proficiency in a second language, learners need to be able to use appropriately the many different types of formulaic sequences that exist in language (Siyanova-Chanturia & Pellicer-Sánchez 2019). As Nation & Webb (2011) argue, 'knowledge of multiword units can be approached in the same way as knowledge of single words' (p.189). The depth of the knowledge of a particular set of sequences can also be explored using a multi-component approach. Nation & Webb (2011) provide a comprehensive list of the components involved in knowledge of multiword units (adapted from Nation 2001, for single words), including several aspects of the form, meaning and use of the sequences, both at the receptive and productive level. Given the prominent role that the study of formulaic language has in vocabulary research, the study by Schmitt could be replicated using the same methodology but changing the focus from individual words to formulaic sequences. This replication approach would allow researchers to compare the learning of formulaic sequences with learning gains for single words in the original study, a type of comparison that is scarce in vocabulary learning research (Pellicer-Sánchez, in press). In fact, Schmitt (1998) initially included collocation as one of the aspects to measure but, as he explains, the measurement instrument needed further development and test results were eliminated. Given the nature of formulaic language and the many different types of sequences that exists in language, it is unlikely that one close replication study could explore the development of more than one or two types of formulaic sequences. Researchers should choose the specific type of formulaic sequence to include in the study. Some of the currently available lists of formulaic sequences might prove useful in the selection of target items. For example, researchers might decide to explore the development of polysemous phrasal verbs, for which the use of the PHaVE List (Garnier & Schmitt 2014) would be convenient.

3. Replication Study 2: Webb & Chang (2012)

3.1. Background to the study

Webb & Chang's (2012) study aimed at shedding light on the amount of words that are learned over four years of instruction. Most research studies exploring changes in learners' vocabulary size have focused on the investigation of relatively short interventions (e.g. Cobb & Horst 2001). As stated in the introduction to this paper, very few studies have looked at second language learners' vocabulary growth over longer periods of time, and the existing studies have yielded conflicting results. The studies by Milton (2006, 2008) did show improvements in vocabulary size, whereas the study by Clark & Ishida (2005) failed to show a difference in learners' vocabulary size over the course of one semester. In addition, the few available studies have employed a cross-sectional design. The study by Webb & Chang (2012) is the only one that looked at the vocabulary growth of the same group of learners. As the researchers argue, previous research has shown that English as a Foreign Language (EFL) students often fail to learn even the most frequent words in English after several years of study (e.g. Nurweni & Read 1999), which points to the inefficiency of vocabulary instruction in some EFL contexts. Therefore, in order to inform language pedagogy we need to have a better understanding of the amount and type of words that are learned at different points in the learning process. To this end, Webb & Chang (2012) examined the number of words that EFL learners in Taiwan learned over four years of instruction, as well as the frequency profile of those words.

Participants in the study were 166 EFL learners in Taiwan. Participants were initially 15-16 years old and 20-21 at the point of the last data collection session. They were from six intact classes in a vocational senior high school. Importantly, the six classes differed in the amount of exposure to English they received. The six classes were divided into three groups

according to the amount of English instruction. Classes belonging to Group A were oriented towards English language and literature and received a larger amount of English instruction (10-22 hours per week across the four years), whereas classes in Groups B and C took English language learning as one of the different study requirements and therefore the amount of instruction was less (0-4 hours per week across the four years). Participants' vocabulary size was measured using a bilingual version of the Vocabulary Levels Test (VLT) (Nation 1990; Schmitt, Schmitt, & Clapham 2001), including five levels: 1,000, 2,000, 3,000, 5,000 and AWL. The first two levels were taken from the bilingual version of the VLT developed by Nation & Wang (available at https://www.victoria.ac.nz/lals/about/staff/paul-nation) and the remaining levels were created for the purposes of this study, translating the items from the original monolingual version. The same test was used every year but the order of items in each level changed between the years. In the analysis Webb & Chang (2012) looked at the scores per level and the sum of scores across the five levels.

Results of the study revealed a gradual increase in vocabulary knowledge each year. Interesting group differences in the size of that increase emerged. The greatest increase was experienced by Group A with 12.91 points, which represents an improvement of 430 words a year (on the assumption that one item in the VLT represents knowledge of 33.3 words). The amount of instruction that the groups received played a significant role in how much learners' vocabulary knowledge grew. The average increase in total scores for Group A, which received a considerably higher amount of instruction, was 9.0 points, whereas for Groups B and C the increase was 4.68 and 4.83 respectively. Webb & Chang (2012) also looked at increases within each of the frequency levels they assessed. The increase within each level ranged from 2.79 at the 5,000 level to 5.17 points at the AWL level. The greatest increase was at the 2,000 and AWL levels for the three groups. The increase at the other three levels (1,000, 3,000, and 5,000) was similar. Importantly, results also showed that by the end of the

study 73% of the participants in Group A had mastered the 1,000 and 33% had mastered the 2,000 level (based on the argument made by Schmitt, Schmitt, & Clapham 2001, that a score of 26 in any one level indicated mastery of the level). On the contrary, only 34% and 29% of students in Groups B and C respectively had mastered the 1,000 level. Only three participants in Groups B and C (7% and 5% respectively) mastered the 2,000 level. Overall, results of this study showed that there was considerable variation in the vocabulary growth between groups of participants who received different amounts of instruction and between years of study. The minimum increase experienced by some of the participants (18 and 76 words in a year for two groups) clearly points towards the need for a more principled approach to vocabulary instruction in this context and for an institutional vocabulary learning plan. Interestingly, even after nine years of English instruction, a considerable percentage of participants failed to master the most frequent levels, which led the authors to conclude that a greater focus on the higher frequency vocabulary should be placed.

3.2. Approaches to replication

The results of Webb & Chang (2012) clearly showed that the amount and type of instruction provided in certain contexts might not be sufficient to support learners in acquiring the large vocabulary learning targets that they are required to operate in English. Given the important implications that this has for curriculum design and institutional planning, it would be useful to replicate the study to substantiate the findings and assess their generalisability. As Webb & Chang (2012) argue, because of the many factors that affect the amount of learning over time, 'L2 vocabulary growth in one situation may be quite different from another' (p. 114).

Therefore, a first approach to approximate replication would involve a series of modifications in the design to explore vocabulary growth in other EFL contexts and assess the generalisability of the original findings. Educational contexts differ in the amount of

exposure that learners get inside the classroom and, importantly, outside the classroom context. This certainly has an important effect on the number of words that learners know and that they learn every year. Higher vocabulary gains than those found by Webb & Chang (2012) could be found in other contexts with more out-of-class exposure to English language. Thus, the first modification would be to conduct the study in an EFL context where learners are usually exposed to English outside the classroom context. The study by Webb & Chang (2012) provided evidence for the important role that amount of in-class exposure had on learners' vocabulary development. However, the effect of out-of-class exposure was not investigated, possibly because of the little exposure that these learners had to English language outside the classroom. Recent investigations have provided evidence for the important role that out-of-class exposure has in the development of learners' lexical knowledge (Schmitt & Redwood 2011; Lindgren & Muñoz 2013; González-Fernández & Schmitt 2015). A recent study by Peters' (2018), also conducted in an EFL context with secondary school students, found that both amount of instruction and out-of-class exposure to English language had an important effect on learners' vocabulary knowledge, with out-of-class exposure having a larger effect. A second modification would involve the use of instruments to elicit information about learners' out-of-class exposure to English. Following the methodology employed in Peters's (2018) study, researchers could add a questionnaire about the amount and nature of out-of-class exposure, such as the European Survey of Language Competences (ESLC) (European Commission, 2012), where learners have to indicate how often they are exposed to English language songs, (subtitled) TV or movies, computer games, websites, books, and magazines. As Peters (2018) noted, the design of the study did not allow her to draw any conclusions as to the longitudinal effect of out-of-class exposure on learners' vocabulary knowledge. The approximate replication suggested here would allow researchers to explore not only learners' vocabulary growth over four years, but

also the longitudinal effect of in-class and out-of-class exposure on vocabulary development. The study by Webb & Chang (2012) used a bilingual version of the VLT. We should note, however, that conducting the study in other EFL contexts might mean using some other bilingual versions or the monolingual version in cases where the bilingual version is not available.

A second approach to approximate replication that also focuses on modifications to the learning context would be to replicate the study in the context of Content and Language Integrated Learning (CLIL), a prominent educational trend in many European contexts. The development of vocabulary is a key concern for CLIL researchers and teachers (e.g. Pérez-Vidal 2007; Costa 2012). In the original study by Webb & Chang (2012), the three groups differed in the number of hours of instruction that they received but were also attending programmes that had a very different nature. A first modification might therefore involve conducting the study in the context of CLIL, which would be a different operationalisation of in-class exposure. Secondly, comparing CLIL and non-CLIL classes in the same context, which follow the same curriculum and only differ in the medium of instruction of certain subject, would allow us to have a clearer idea of the contribution of in-class exposure to vocabulary growth. A third methodological modification might involve administering a language exposure questionnaire, along the lines of what I suggested above, in order to explore the effect of both in-class and out-of-class exposure. Another interesting context for a further approximate replication would be the SA context, as in the study by Fitzpatrick (2012) reviewed above. In SA contexts the amount of out-of-class exposure is expected to be larger than in an EFL context. Moreover, the type of exposure is also expected to be different, with more opportunities for oral interaction. The comparison of results of these approximate replications conducted in different contexts to the results of the original study would shed light on the effect of context and exposure on the development of vocabulary knowledge.

A final approach to approximate replication would be to modify the vocabulary measures used in order to provide a more detailed picture of vocabulary growth. The original study used a bilingual version of the VLT, which was deemed appropriate for the low-proficiency participants in the study. It measures vocabulary knowledge at different frequency levels. Thus, it provides information about the amount of words they know at each level, but does not provide an overall figure of vocabulary size (Schmitt 2010). Vocabulary growth could also be measured using the Vocabulary Size Test (VST) (Nation & Gu 2007), which is indeed intended to test overall vocabulary size and some bilingual versions are also available. The VLT and the VST could be both administered at different points in a year, allowing us to see which one is more effective in capturing changes in vocabulary knowledge over time. The VLT is a receptive measure. Thus, a further modification in the measures used would involve adding a measure of productive vocabulary knowledge such as the Productive Levels Test (Laufer & Nation 1999). This would allow us to look at the growth of both receptive and productive vocabulary knowledge.

4. Conclusion

The gradual and incremental nature of vocabulary knowledge implies the need to explore vocabulary development over a long period of time, both in terms of the size of learners' vocabulary and the depth of that knowledge. However, longitudinal investigations of vocabulary growth are still scarce. Regarding depth of vocabulary knowledge, the very few studies that have attempted to look at the development of different word knowledge components have mainly looked at the development of one learner (e.g. Churchill 2008; Bell 2009; Fitzpatrick 2012) or have looked at a very small number of components (e.g. Schmitt & Meara 1997). Schmitt (1998) is the only one that looked at the acquisition of a range of word knowledge aspects with a small group of learners. Concerning vocabulary size, the existing studies looking at the development of learners' quantity of vocabulary knowledge

have mainly employed cross-sectional designs (e.g. Milton 2006; Henriksen 2008; Milton 2008) or have followed learners for a relatively short time (e.g. Milton & Meara 1995). The study by Webb & Chang (2012) remains the only one that has examined learners' vocabulary growth over several years. Because of the uniqueness of these two studies, their important contributions to our understanding of how vocabulary develops over time, together with the scarcity of longitudinal investigations of vocabulary learning, the need for replicating these studies is warranted. The results of the replications here suggested would have important implications for vocabulary teaching practice, as well as for the building of an unaccounted theory of vocabulary learning.

References

Anderson, R. C. & P. Freebody (1981). Vocabulary knowledge. In J. Guthrie (ed.), *Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association, 77–117.

Avila, E. & M. Sadoski (1996). Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning* 46.3, 379–395.

Bell, H. (2009). The messy little details: A longitudinal case study of the emerging lexicon. In T. Fitzpatrick & A. Barfield (eds.), *Lexical processing in second language learners.* Bristol, UK: Multilingual Matters, 111–127.

Churchill, E. (2008). A dynamic systems account of learning a word: From ecology to form relations. *Applied Linguistics* 29.3, 339–358.

Clark, M. K. & S. Ishida (2005). Vocabulary knowledge differences between placed and promoted EAP students. *Journal of English for Academic Purposes* 4.3, 225–238.

Cobb, T. & M. Horst (2001). Reading academic English: Carrying learners across the lexical threshold. In J. Flowerdew & M. Peacock (eds.), *The English for academic purposes curriculum*. Cambridge: Cambridge University Press, 315–329.

Costa, F. (2012). Focus on form in ICLHE lectures in Italy: Evidence from English-medium science lectures by native speakers of Italian. *AILA Review* 25, 30–47.

Dang, T.N.Y. & S. Webb (2014). The lexical profile of academic spoken English. *English for Specific Purposes* 33, 66-76.

European Commission (2012). *First European survey on language competences.*

Garnier, M. & N. Schmitt (2014). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research* 19.6, 645–666. https://doi.org/10.1177/1362168814559798

González Fernández, B. & N. Schmitt (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *ITL International Journal of Applied Linguistics* 166.1, 94–126. https://doi.org/10.1075/itl.166.1.03fer

Goulden, R., P. Nation & J. Read (1990). How large can a receptive vocabulary be? *Applied Linguistics* 11.4, 341–363.

Huckin, T., M. Haynes & J. Coady (1993). *Second language reading and vocabulary learning.* Norwood, NJ: Ablex.

Fitzpatrick, T. (2012). Tracking the changes: Vocabulary acquisition in the study abroad context. *The Language Learning Journal* 40.1, 81–98.

Francis, W. N. & H. Kucera (1982). *Frequency analysis of English usage*. Boston: Houghton Mifflin.

Henriksen, B. (2008). Declarative lexical knowledge. In D. Albrechtsen, K. Haastrup & B. Henriksen, (eds.), *Vocabulary and writing in a first and second language*. Basingstoke: Palgrave Macmillan, 22–66.

Laufer, B. & P. Nation (1999). A vocabulary size test of controlled productive ability. *Language Testing* 16, 36–55.

Lindgren, E. & C. Muñoz (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism* 10.1, 105–129. https://doi.org/10.1080/14790718.2012.679275

Milton, J. (2006). Language Lite: Learning French vocabulary in school. *Journal of French Language Studies* 16, 187–205.

Milton, J. (2008). French vocabulary breadth among learners in the British school and university system: comparing knowledge over time. *French Language Studies* 18, 333–348. doi:10.1017/S0959269508003487

Milton, J. & T. Fitzpatrick (2014). Introduction: Deconstructing vocabulary knowledge. In J. Milton & T. Fitzpatrick (eds.), *Dimensions of vocabulary knowledge*. London: Palgrave Macmillan, 1–13.

Milton, J. & P. Meara (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL International Review of Applied Linguistics* 107-108, 17–34.

Nagy, W. E., I. N. Diakidoy & R. C. Anderson (1993). The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives. *Journal of Reading Behavior* 25.2, 155–170.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Rowley, Mass.: Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63.1, 59–82.

Nation, P. & P. Y. Gu (2007). *Focus on vocabulary.* Sydney: National Centre for English Language Teaching and Research.

Nation, I.S.P. & S. Webb (2011). *Researching and analysing vocabulary*. Boston, MA: Heinle, Cengage Learning.

Nurweni A. & J. Read (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes* 18.2, 161–75.

Pellicer-Sánchez, A. (in press). Learning single words vs. learning multiword units. In S. Webb (ed.), *The Routledge handbook of vocabulary studies*. New York: Routledge.

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics* 169.1, 142–168.

Pérez-Vidal, C. (2007). The need for focus on form in content and language integrated approaches: an exploratory study. In F. Lorenzo, S. Casal, V. de Alba & P. Moore (eds.), *Models and practices in CLIL. Revista Española de lingüística aplicada (RESLA).* Logroño: Asociación Española de Lingüística Aplicada, 39–53.

Pigada, M. & N. Schmitt (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language* 18.1, 1–28.

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

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: Longitudinal study. *Language Learning* 48.2, 281–317.

Schmitt, N. (2004). *Formulaic sequences*. Amsterdam: John Benjamins.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* London: Palgrave Macmillan.

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

Schmitt, N. & S. Redwood (2011). Learner knowledge of phrasal verbs. A corpus-informed study. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (eds.), *A taste for corpora: In honour of Sylviane Granger*. Amsterdam: John Benjamins, 173–209. https://doi.org/10.1075/scl.45.12sch

Schmitt N., D. Schmitt & C. Clapham (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18.1, 55–88.

Siyanova-Chanturia, A. & A. Pellicer-Sánchez (2019). Formulaic Language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicer- Sánchez (eds.), *Understanding Formulaic Language: A second language acquisition perspective*. London: Routledge, 1-17.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition* 27, 33–52.

Webb, S. & A. Chang (2012). Second language vocabulary growth. *RELC Journal* 43.1, 113–128.

Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence, and classroom applications.* London and New York, NY: Continuum.

Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. New York, NY: Bloomsbury.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic language: Pushing the boundaries.* Oxford: Oxford University Press.

Xue, G. & P. Nation (1984). A university word list. *Language Learning and Communication* 3.2, 215–229.